

RESULTADOS DE PRUEBAS DE ENSAYO Y SELECCION AL EVALUAR
LOS MISMOS OBJETIVOS DE UN CURSO INTRODUCTORIO DE
MATEMATICA SUPERIOR

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Educación

RESULTADOS DE PRUEBAS DE ENSAYO Y SELECCION AL EVALUAR.
LOS MISMOS OBJETIVOS DE UN CURSO INTRODUCTORIO DE
MATEMATICA SUPERIOR

UNIVERSIDAD DEL VALLE DE GUATEMALA

BALTAZAR ROLANDO NUÑEZ LAGOS

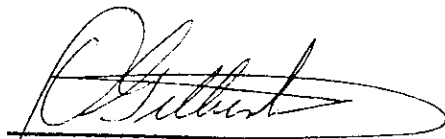
Trabajo de investigación presentado para optar al Grado
Académico de Maestría en Medición, Evaluación e
Investigación Educativas

Guatemala

1982

Vo. Bo.

(f)



Doctor Otto E. Gilbert A.
Asesor

Fecha de aprobación: 11 de octubre de 1982

A mis padres con gratitud y afecto

A mi esposa e hijos, murallas de fortaleza moral y espiritual en mi vida de trabajo y estudio

A mis hermanos

A mis amigos

AGRADECIMIENTO

En la realización de una meta profesional se conjugan una serie de situaciones y de personas que permiten mediante su estímulo y apoyo, alcanzarla.

Deseo dejar constancia de mi gratitud a todos los miembros del personal docente del Departamento de Matemática de la Escuela Superior del Profesorado "Francisco Morazán".

Aprovecho la ocasión para agradecer profundamente la confianza depositada en mi persona, por el Ingeniero Benjamín Henríquez, Jefe del Departamento de Matemática de la Universidad Nacional Autónoma de Honduras, extensivo para mis compañeros que con un gesto de amistad, espontáneamente me brindaron su apoyo.

Efusivamente quiero expresar mis sinceros agradecimientos al Doctor Otto E. Gilbert A., por su valiosa colaboración y orientación brindada con invariable cordialidad y gentileza en la asesoría del presente trabajo.

Mis agradecimientos: a los catedráticos del Programa de Maestría por sus acertadas enseñanzas y espíritu de amistad; a mis compañeros de la Maestría y amigos guatemaltecos que colaboraron y apoyaron en diversas ocasiones y circunstancias.

Para todos, mis sincero agradecimiento.

ABSTRACTO

El deficiente rendimiento del estudiante en los cursos de matemática en todos los niveles del sistema educativo hondureño, particularmente en el nivel superior, hace necesario indagar sus posibles causas.

El objetivo central del presente trabajo es evidenciar si hay diferencia estadísticamente significativa entre los resultados obtenidos por el estudiante, al aplicarle una prueba estructurada del tipo de selección múltiple y otra de ensayo, en la evaluación de los mismos objetivos de un programa introductorio de matemática a nivel superior.

Se utilizó el método de investigación "Experimento de campo", con una población constituida por los alumnos de primer ingreso de la Escuela Superior del Profesorado "Francisco Morazán" que cursaron la asignatura Matemática I (MAT 101) en el primer semestre del año 1982. Como instrumentos se utilizó:

1. Prueba de base estructurada formada por ítemes de selección múltiple.
2. Prueba de base no estructurada formada por ítemes de ensayo.
3. Además se elaboró la tabla de especificaciones respectiva.

Se llevó a cabo un estudio piloto con el propósito de depurar los instrumentos, para medir la unidad "conjuntos de los números reales", posteriormente se aplicaron a la muestra de 22 estudiantes de la población seleccionada.

El estadístico empleado en la verificación de las hipótesis fue un contraste de diferencias entre medias mediante una "t" de Student, obteniéndose un valor $t_{\bar{x}_A - \bar{x}_B} = 0.9$; al compararlo con la "t" crítica = 2.080 obtenida en la tabla a un nivel $p_{\alpha} = 0.05$ y 21 grado de libertad, se estableció la siguiente conclusión: Probablemente con un nivel de significación de error alfa igual a 5 por ciento ($p_{\alpha} = 0.05$) la hipótesis nula, H_0 : "No existe diferencia estadísticamente significativa entre el rendimiento obtenido por los alumnos al aplicarles pruebas de selección múltiple y pruebas de ensayo, para evaluar los mismos objetivos de una unidad del programa de Matemática I (MM101)", no se rechaza. En base a los resultados obtenidos se recomienda:

- a. Extender este tipo de estudio, a las demás áreas de formación de la Escuela Superior y de la Universidad Nacional.
- b. Tratar que en futuros estudios similares la prueba de selección múltiple y la prueba de ensayo formen parte de los exámenes regulares de aprovechamiento, en la evaluación de los participantes.

CONTENIDO

	Páginas
AGRADECIMIENTO	ix
I. INTRODUCCION	1
A. Antecedentes	1
B. Situación actual del problema	3
C. Justificación	4
D. Proyección	5
E. Objetivos	5
II. FUNDAMENTACION TEORICA	7
A. Pruebas de base no estructurada	10
B. Pruebas de base estructurada	16
1. Concepto	16
2. Características	18
3. Ventajas	18
4. Desventajas	19
5. Tipos de ítemes	20
C. Rendimiento	23
D. Investigaciones similares	26
E. Comentario sobre la revisión bibliográfica	30
III. METODOLOGIA	33
A. Problema	33
B. Hipótesis	33
C. Variables	33
D. Diseño experimental	35

	Páginas
E. Instrumentos y mediciones	35
F. Población y muestra	38
G. Pruebas estadísticas	39
H. Procedimiento	39
IV. RESULTADOS	43
A. Cuadros informativos	43
B. Prueba estadística	46
C. Interpretación de resultados	47
V. DISCUSION DE LOS RESULTADOS	49
A. Conclusiones	49
B. Limitaciones	50
C. Recomendaciones	51
VI. BIBLIOGRAFIA	53
ANEXOS	
A. Tabla de especificaciones	55
B. Prueba estructurada	57
C. Prueba no estructurada	61

I. INTRODUCCION

Es innegable el papel que en la sociedad actual está desempeñando la matemática, dada la gran importancia a la cuantificación, implicando su reconocimiento como herramienta esencial del pensamiento y como medio para interpretar y comprender el mundo en que vivimos. Desde este ángulo es muy importante evaluar todos los factores que puedan afectar su aprendizaje, entre ellos, la selección y elaboración de instrumentos adecuados para evaluar el rendimiento de los estudiantes.

A. Antecedentes

En Honduras, a nivel de la educación superior, el profesor ha tenido libre criterio para medir el rendimiento del alumno. A partir de 1970 se adopta, por decisión de los Coordinadores de Area, las pruebas estructuradas, especialmente del tipo de selección múltiple, conjuntamente con las pruebas de ensayo, en la elaboración de pruebas de matemática. Esto se hace con el propósito de unificar objetivos y contenidos, para medir grupos numerosos de diferentes carreras que siguen un mismo programa de estudio, estableciéndose las coordinaciones de área para asegurar el buen funcionamiento del plan. A pesar de la reestructuración de los programas en términos de objetivos de comportamiento; el uso de libro de texto y de guías complementarias, para unificar contenidos y metodología, el porcentaje de reprobación actual es de 60 por ciento. No se ha realizado un trabajo serio para analizar las variadas causas de la reprobación. Existe preocupación por encontrar una posible solución al problema. En este trabajo se estudia uno de los posibles factores: la elaboración de prue-

bas y sus efectos sobre el rendimiento del alumno. Una tendencia creciente en el campo de la educación es hacer de la medición y evaluación en el aula, un medio de promoción de más y mejores aprendizajes. La medición y la evaluación son procesos íntimamente relacionados, pero no se deben confundir. La medición consiste en descubrir cuáles son las características de alguien o algo y asignarles símbolos, a menudo numéricos.

La evaluación del aprendizaje significa juzgar la importancia, el valor o mérito de alguien o algo, con el propósito de verificar los aciertos y dificultades para la toma de decisiones. Actualmente la evaluación de los productos del aprendizaje es uno de los temas fundamentales de la educación. Terrón Estrada (1978:7) dice:

"La evaluación es un proceso sistemático, continuo e integral, que estableciendo juicios de valor permite tomar decisiones".

Burton (1970) y Gronlund (1973) señalan como metas de la evaluación:

1. Clarificar los objetivos del aprendizaje.
2. Conocer y comprender al que aprende.
3. Motivar al alumno para el aprendizaje.

El profesor como diseñador, orientador y verificador del aprendizaje, debe promover el acercamiento, la comunicación interpersonal y el estímulo verbal como esencia del proceso educativo. Él intenta convertir la educación de la población en educación de individuos. Lamentablemente los grupos numerosos de alumnos en los cursos de preparatoria en el nivel

superior, dificultan el logro de estos objetivos.

Es por éso que se considera que la elaboración de pruebas y su utilización deben dejar de ser el procedimiento tradicional de asignar una calificación al alumno. Las pruebas deben transformarse en instrumentos que permitan generar un espíritu activo y social dentro de las aulas; es decir, han de ser medios para una adecuada motivación, reorientación de objetivos, o replanteamiento de la organización total del curso y fuentes de retroalimentación de alumnos, profesores y de la administración escolar.

B. Situación actual del problema

No se pretende sostener que en el proceso de enseñanza-aprendizaje de la matemática la medición sea inadecuada; se enfatiza el hecho de ser insuficiente para verificar el logro de los objetivos que un educando debe alcanzar dentro del contexto de la educación integral. Actualmente la evaluación se orienta más hacia las normas (la que el alumno domina comparado con otros miembros del grupo) y no a criterios (relación de un alumno con lo que verdaderamente domina de los objetivos propuestos).

En el campo de la matemática a nivel superior, el profesor fundamentalmente mide los conocimientos adquiridos por el alumno, mediante la elaboración de pruebas, cuya estructuración en general es de un 50 ó 60 por ciento de selección múltiple y 50 ó 40 por ciento respectivamente, de ensayo, consistente en situaciones -problema, que el alumno debe resolver.

Al elaborar una prueba de matemática, el profesor no siempre considera:

1. La estructuración de una tabla de especificaciones.
2. La formulación adecuada de las preguntas o ítemes respectivos.
3. La estimación del tiempo necesario para que el estudiante la conteste.
4. La evaluación de la prueba en un tiempo prudencial, que permita la discusión grupal e individual de la misma.
5. Criterios de evaluación uniformes en la parte de la prueba de ensayo, que modere la influencia del prejuicio personal, efecto del halo y el error lógico; aspectos que afectan la confiabilidad del instrumento. En la educación superior, con raras excepciones, el profesor no le da importancia a los aspectos científicos y didácticos de la instrucción. Esto produce un enfoque antipedagógico de la enseñanza, que va desde la formulación de objetivos hasta la evaluación de los resultados obtenidos.

El típico profesor universitario ve con indiferencia y franco antagonismo los aspectos pedagógicos, lo cual constituye una actitud negativa que debe cambiarse.

C. Justificación

El deficiente rendimiento del estudiante en los cursos de matemática en todos los niveles del sistema educativo, pero particularmente en el caso que nos ocupa, del nivel superior, hace necesario indagar cuáles son

sus causas. Con ese propósito, en esta investigación, se midió el rendimiento del alumno en matemática por medio de dos instrumentos: una prueba construida con sólo ítemes de ensayo; y otra preparada con ítemes de selección múltiple, con cuatro alternativas; evaluando ambas los mismos objetivos de un programa de matemática. Con base en los resultados se estudió su eficiencia y sus limitaciones, se diagnosticó posibles fallas y se obtuvo conclusiones válidas y confiables acerca del logro de los objetivos alcanzados por el alumno.

D. Proyección

Los resultados de este estudio podrán ser utilizados por:

1. Los centros de educación superior, especialmente por los departamentos de matemática de la Universidad Nacional y de la Escuela Superior del Profesorado, como información valiosa en la toma de decisiones para modificar el enfoque y estrategias de la enseñanza y los instrumentos utilizados para medir el logro de los objetivos propuestos en sus programas de estudio.

2. Por el Ministerio de Educación, como referencia en investigaciones similares del sistema educativo, conducentes a la implementación de la reforma educativa.

E. Objetivos

1. Evidenciar si hay diferencia estadísticamente significativa entre los resultados, al aplicar una prueba estructurada del tipo de selección múltiple y otra de ensayo, en la evaluación de los mismos obje-

tivos de un programa de matemática.

2. Hacer conciencia sobre el verdadero valor retroalimentador que tienen las pruebas de aprovechamiento en el aprendizaje.
3. Motivar a los profesionales de la Educación para que realicen investigación en este campo u otros similares.

II. FUNDAMENTACION TEORICA

La evaluación del rendimiento es una de las actividades primordiales de la labor educativa que está basada, en su mayor parte, en las pruebas efectuadas en el aula. Corresponde al profesor verificar los resultados del proceso educativo, cuantificando el progreso mostrado por sus alumnos.

Considerando que los objetivos que se pretenden alcanzar con determinados contenidos, condicionan el tipo de prueba a utilizar, las pruebas construídas por el profesor se clasifican en dos tipos:

1. Estructuradas: donde el alumno elige entre varias alternativas, marcando la que él considera más acertada, como en el caso de las de selección múltiple.
2. No estructuradas: donde el alumno construye sus respuestas de acuerdo a los problemas dados por el profesor.

Estas pruebas, bien elaboradas, facilitarán el logro de objetivos importantes en la formación del alumno, evitando que su calificación sólo signifique propósitos secundarios que él quiere alcanzar porque así conviene a sus intereses inmediatos.

Actualmente, como actividades importantes para el alumno en el periodo lectivo en lo que respecta a medición y evaluación se destacan: la ejecución de pruebas; la devolución de las mismas y el conocimiento de la calificación que ha merecido en una tarea y otras actividades similares.

Históricamente se puede señalar el predominio de algunas concepciones evaluativas que se han utilizado simultáneamente, entre ellas; según Nilo (1973:5-6).

"Las concepciones inestructuradas: se refieren a la evaluación como juicio del que sabe, un juicio del experto".

"La evaluación considerada como medición, surge como la parte científica de la educación. La objetividad del instrumento, opuesta a la subjetividad del experto. Así, la evaluación se convierte en la técnica de elaboración de instrumentos y de interpretación de sus resultados".

Para evaluar los cursos de matemática en el nivel superior generalmente se enfatiza la medición de los resultados cognoscitivos, reflejados en el rendimiento alcanzado en pruebas de conocimiento, este procedimiento provoca frustraciones desde sus inicios.

Según Bloom (1975:22), desde el punto de vista de los profesores puede señalarse que:

"Tal como se usan en la escuela las pruebas y otras formas de evaluación, se observa que contribuyen poco al mejoramiento de la enseñanza-aprendizaje y raras veces sirven para asegurar que todos (o casi todos) aprendan lo que el sistema escolar considera tareas y metas importantes del proceso educativo".

El problema capital que se plantea a los profesores en la elaboración de una prueba, es la de dar respuesta a las preguntas: ¿Qué medir? y ¿Cómo medir? La primera, tiene relación con una concepción válida de la naturaleza de los objetivos generales y específicos de la enseñanza y sus logros evidenciados por el rendimiento educativo; la segunda, con la selección de la prueba y el tipo de ítem más adecuado y eficiente para medir los objetivos propuestos.

Gronlund (1973:600), presenta las categorías básicas de objetivos en los dominios: cognoscitivo, afectivo y psicomotor, que últimamente han servido de base en la construcción de ítemes de rendimiento.

En educación se ha hecho intentos de ordenar las diferentes pruebas e instrumentos empleados con mayor frecuencia. Los criterios de clasificación no siempre han definido claramente la naturaleza de los objetivos en un determinado ordenamiento universal. Corrientemente se observa que las clasificaciones de pruebas enfatizan el campo cognoscitivo.

Según Thorndike y Hagen (1980:54), existen diferentes tipos de ítemes que el profesor puede utilizar al elaborar una prueba, clasificándolos en dos categorías principales:

El tipo conforme al cual el alumno:

1. Aporta su propia respuesta, tales como:
 - a. La pregunta de ensayo o respuesta libre, que requiere una respuesta global del alumno.
 - b. La respuesta breve, que requiere a lo más dos oraciones para ser contestada.
 - c. La respuesta de completación, que necesita una palabra o una frase para integrar la respuesta.
2. Elige su respuesta, dentro de varias alternativas proporcionadas por el profesor, citándose los ítemes de:

- a. Respuestas optativas, como el verdadero y falso.
- b. Elección múltiple.
- c. De igualamiento o apareamiento.

A. Prueba de base no estructurada: Ensayo

Se le define como preguntas o temas dados por el profesor, ante los cuales el estudiante tiene que producir sus propias respuestas; debe demostrar a través de su organización, su originalidad, su integración y su expresión de ideas para responder a un planteamiento.

Thorndike y Hagen (1980:88) expresan respecto a la pregunta de ensayo lo siguiente:

"El valor principal de la pregunta de ensayo estriba en lo bien que se presta para medir las capacidades del alumno para organizar, integrar y sintetizar sus conocimientos. Para emplear su información a fin de resolver problemas nuevos, o de ser original, o innovador en su manera de abordar problemas. Para realizar este potencial se necesita que cada pregunta esté cuidadosamente expresada, de manera que el alumno revele todas estas clases de capacidades. El simple dar forma de ensayo a una pregunta no garantiza automáticamente que se podrán estimar estas capacidades".

En el campo de la matemática, "The National Council of Teachers of Mathematics" (1961:61), explica y ejemplifica lo que considera un ítem de ensayo; en la forma siguiente:

"Essay items require that the student produce answer. He must select and organize the ideas he wishes to include in an answer and express these ideas in writing. This, "prove that $\sqrt{2}$ is an irrational number" is an item that tests communication skill as well as an understanding of a deductive proof".

Traducido al español el párrafo anterior por F. Jiménez (1979:15)

"Los ítemes de ensayo requieren que el estudiante construya la respuesta. Debe seleccionar y organizar las ideas que desea incluir en la respuesta y expresarlas en forma escrita. Así; "pruebe que $\sqrt{2}$ es un número irracional", es un ítem que mide habilidad de comunicación así como también la comprensión de una prueba deductiva".

1. Características de los ítemes de ensayo

a. Permiten al alumno:

- 1) Libertad para responder
- 2) Demostrar su originalidad y creatividad
- 3) Medir sus actitudes y métodos de pensamiento
- 4) Organizar sus propias respuestas

b. Generan productos complejos

c. Son difíciles de calificar

d. No existe un modelo uniforme para evaluar la respuesta

2. Ventajas

a. Proporcionan evidencia directa de los productos complejos del aprendizaje que no pueden medirse por otros medios.

b. Permiten la integración y la aplicación de las habilidades del pensamiento y de la resolución de problemas

- c. Son aparentemente fáciles de construir
- d. Se adaptan a la mayoría de las asignaturas
- e. Obligan al estudiante a fomentar buenos hábitos de estudio, a estudiar con profundidad y a expresarse lingüísticamente

bien

- f. Estimula la capacidad de razonamiento y la creatividad
- g. Proporcionan economía en cuanto a gastos de materiales.

3. Desventajas

- a. Validez y confiabilidad variable
- b. Se prestan a la evaluación subjetiva
- c. Dificiles de calificar
- d. Requieren de personas expertas para revisarlas y calificarlas
- e. El muestreo de contenidos es generalmente muy limitado.

4. Tipos de ítemes en las pruebas de ensayo

- a. De respuesta amplia o extensa: Son preguntas que se plantean para que el estudiante demuestre con libertad sus capacidades de enfocar, organizar y desarrollar la respuesta.

Sirven para evaluar habilidades en forma integrada, tales como las capacidades de seleccionar, organizar, integrar y evaluar ideas.

Son valiosas tanto para la enseñanza como para la evaluación, aunque la gran mayoría de las veces, al plantearlas, no se redactan para que el estudiante exhiba su capacidad y su originalidad.

Muchas veces al expresar las ideas, el estudiante no sólo manifiesta las suyas, sino que también responde en función de lo que el docente desea que responda.

Al producir la respuesta el alumno utiliza el vocabulario que maneja, su propio estilo de redacción y su propia caligrafía. Estos aspectos afectan la calificación, cuando el docente no ha establecido claramente los criterios a considerar.

b. Ensayo de respuestas restringidas. Son preguntas que se le presentan al alumno para que responda en forma limitada, ya que se le definen qué aspectos debe tratar, se le indica la naturaleza de los límites descados de la respuesta.

Se utilizan para medir productos del aprendizaje que requieren de interpretación y aplicación de los datos en un área específica.

Sirven para medir el reconocimiento de la relación causa-efecto, la aplicación de principios y para explicar métodos y procedimientos.

Se requiere menos tiempo para calificar las respuestas que para las

de respuesta amplia.

5. Proceso de construcción de las pruebas tipo ensayo. Para hacerlas más eficientes es necesario:

- a. Definir claramente los objetivos a evaluar.
- b. Determinar el peso (o valor relativo) a asignarle en relación al contenido o nivel del objetivo y al proceso mental explorado.
- c. Determinar la adecuación de la misma en relación con los objetivos a evaluar.
- d. Elaborar una tabla de especificaciones.
- e. Redactar los ítemes y las instrucciones
- f. Informar a los estudiantes, con anticipación, aspectos relacionados con el tipo de prueba (objetivos, materiales, duración, fecha y lugar).
- g. Elaborar una clave de respuesta correcta para cada ítem.

6. Recomendaciones para la elaboración de pruebas tipo ensayo

- a. Elaborar el número de preguntas posibles que evalúen completamente los objetivos.
- b. Examinar con preguntas de tipo ensayo, cuando los objetivos

sean adecuados.

c. Escribir las instrucciones claras y especificar cuáles preguntas requieren de respuesta restringida y cuáles de respuesta amplia.

d. Revisar los ítemes con otras personas, si es posible.

e. Diseñar previamente criterios de evaluación de las pruebas, con su respectiva escala.

7. Recomendaciones para la administración de las pruebas.

a. Informar con anticipación los objetivos, contenidos, tipos y condiciones de presentación de la prueba.

b. Dar al inicio de la prueba instrucciones y orientaciones precisas que eviten confusiones y permitan rendir más.

c. Presentar las preguntas en orden creciente de dificultad.

d. Informar sobre el valor de cada una de ellas.

e. Impedir el intercambio de información entre los examinados.

f. Evitar la entrada de personas extrañas, durante la aplicación.

8. Métodos de revisión de las pruebas tipo ensayo. Entre éstos se encuentran:

a. Método global o de selección

- 1) El evaluador lee todas las pruebas, una a una para tener idea sobre la calidad de cada respuesta y va formando grupos de pruebas por categorías (deficientes, regulares, buenas y excelentes).
- 2) Releer las pruebas con mayor cuidado a fin de reagruparlas.
- 3) Asignar a cada grupo un intervalo de calificaciones.
- 4) Asignar la calificación a cada prueba.

b. Método analítico

- 1) Determinar los aspectos que se tomarán en cuenta al evaluar cada respuesta y asignarles pesos apropiados.
- 2) Revisar las respuestas dadas por los alumnos en función de un modelo.
- 3) Sumar los porcentajes para formar la calificación, si es que va a ser punteada.
- 4) Indicar los errores, y los aciertos para que éstos sirvan de retroalimentación al estudiante.

B. Pruebas de base estructurada: Selección múltiple

1. Concepto. Las pruebas de selección múltiple son aquellos tipos

de prueba donde el alumno elige sus respuestas entre varias alternativas que aparecen en cada ítem.

Existe consenso entre los autores acerca de lo que es una pregunta o ítem de una prueba de selección, así: Karmel (1974:452) dice lo siguiente:

"El formato de los reactivos de opción múltiple es uno de los más populares y efectivos entre los tests objetivos, consta de dos partes:

1. El encabezado, que establece el problema; y
2. La lista de opciones, una de las cuales tiene que seleccionarse como respuesta.

El encabezado puede expresarse como pregunta o aseveración incompleta".

Lafourcade (1973:106) dice:

"Un ítem de opción múltiple está constituido por una proposición (expresada en forma directa o como una oración incompleta) y una serie de soluciones establecidas en forma de opciones (palabras, frases, símbolos, números), una de las cuales es correcta y las demás actúan como distractores".

Goring (1974:452) expresa lo siguiente:

"Existen dos categorías de selección múltiple:

1. La respuesta correcta y
2. La mejor respuesta.

Ambas categorías admiten una sola opción acertada; sin embargo, el ítem de la categoría "respuesta correcta" trae una alternativa que fría y palpablemente es la única correcta, mientras que un ítem de la categoría "mejor respuesta" ofrece opciones que exigen, una distinción más sutil entre la correcta y los distractores".

2. Características de la prueba de selección. Con el criterio de Thorndike y Hagen (1978:75), las características esenciales de la tarea que pone una prueba objetiva son las que el examinando:

- a. Opera dentro de una situación completamente estructurada.
- b. Elige su respuesta dentro de un número limitado de opciones de respuesta proporcionadas por el que elaboró la prueba.
- c. Responde a cada uno de los ítemes de una gran muestra.
- d. Recibe puntajes por cada respuesta conforme a una clave establecida de antemano.

3. Ventajas

- a. Las pruebas dan medidas comparables de todos los alumnos.
- b. Los resultados de la prueba pueden proporcionar más información diagnóstica acerca del alumno.
- c. El alumno puede responder a todos los aspectos de una área compleja.
- d. El profesor ahorra tiempo en su corrección.
- e. Los diversos calificadores dan iguales puntajes.
- f. Las pruebas pueden ser calificadas por una máquina o cualquier oficinista con algún entrenamiento.
- g. Las pruebas permiten realizar un análisis de ítemes, procedi-

miento mediante el cual es posible determinar cuáles reactivos fueron muy fáciles, cuáles difíciles o ambiguos y cuáles fueron poco válidos para medir los objetivos.

4. Desventajas

- a. La elaboración de ítemes objetivos claros, sin ambigüedades es tarea de considerable exigencia.
- b. La dificultad para encontrar un número suficiente de alternativas homogéneas y razonables.
- c. Las pruebas son anti-económicas en tiempo y materiales cuando hay menos de 20 alumnos.
- d. El riesgo que se corre de que el estudiante adivine la respuesta.

De las desventajas mencionadas se considera oportuno comentar sobre el factor de adivinanza en las respuestas a este tipo de ítem, que: existen problemas en un test con respuestas dadas por el autor, cuando no se aplican fórmulas de corrección por adivinación, ya que cuando ésto ocurre, se incrementa la varianza error y por lo tanto decrece la confiabilidad de la prueba. Según Thorndike y Hagen (1980:137).

"La fórmula que generalmente se emplea para corregir los efectos de la adivinación es: puntuación corregida = $C - \frac{I}{n-1}$

donde C es el número de preguntas contestadas correctamente
 I es el número de preguntas contestadas incorrectamente
 n es el número de opciones de respuesta para un ítem".

Thorndike y Hagen (1980:76) comentan lo siguiente:

"El problema de la adivinanza puede controlarse utilizando ítemes que tengan cuatro o cinco alternativas, poniendo pruebas largas en vez de cortas, elaborando ítemes de dificultad adecuada, proporcionando tiempo suficiente a todos los examinandos para que terminen la prueba y procurando que todos los ítemes tengan opciones erróneas realmente atractivas".

5. Tipos de ítemes. ¿Cuál de los tipos de ítemes es más eficiente para medir el rendimiento?

Para evaluar objetivamente los productos del aprendizaje programado, el profesor necesita utilizar diversidad de ítemes. Es su responsabilidad usar el test adecuado, en las circunstancias apropiadas y para el propósito establecido, considerando cuatro factores teóricos importantes para elegir el tipo de ítem, que según Thorndike y Hagen (1980:78) son:

- "1. La capacidad que el tipo de ítem tenga para provocar la conducta del alumno que estamos tratando de medir con la prueba.
2. El grado de precisión que es necesario alcanzar en los resultados de la prueba, para realizar la finalidad con que se puso la misma.
3. La inmunidad respecto de causas de variación, que no vienen al caso, que se desea tener en los resultados de la prueba.
4. La adecuación a la edad y al nivel de desarrollo de los alumnos que se les aplica la prueba".

En general, no existen reglas definitivas para determinar si la prueba de ensayo o la estructurada es más eficiente, ni tampoco para establecer cuando es más adecuado preferir una u otra. Puede sugerirse que

las pruebas estructuradas tienen prioridad en los niveles de conocimiento correspondientes a: comprensión, aplicación y análisis.

Cuando se quiere que el estudiante proporcione razones, relaciones, conclusiones o que produzca la respuesta adecuada, se recomienda la prueba de ensayo. Si es fundamental que se suministre sólo la respuesta, la más adecuada será la pregunta restringida, bien elaborada. En los niveles de síntesis y evaluación, ambas pruebas tienen un valor limitado. La selección se basa específicamente en lo que el profesor desea medir, en concordancia con el tiempo disponible. Ninguna de estas pruebas, aislada, puede medir satisfactoriamente los logros alcanzados por el alumno.

Al respecto, "The National Council of Teachers of Mathematics (1965: 45-46) dice:

"El tipo de técnicas de respuesta o ítem debe estar determinado en primer lugar por el comportamiento fijado por los objetivos. Desde el punto de vista de la prueba se justifica con mayor razón tener objetivos establecidos en términos de conducta. Los ítemes de ensayo son usados cuando pensamos en la habilidad de comunicarse, esto es, la habilidad para presentar ideas concisas y completas dentro de una organización lógica. Los ítemes objetivos como falso y verdadero, selección múltiple y apareamiento pueden ser usados para el reconocimiento de hechos y relaciones". Traducido por Jiménez Villalobos (1979: 26).

Mediante el siguiente esquema preparado por Gronlund (1974:94), se presenta una comparación sumaria de las pruebas de base estructurada (objetiva) y de las de ensayo.

Cuadro 2.1

Comparación sumaria de los tests objetivos y de ensayo

	Test objetivo	Test de ensayo
Resultados medidos de la taxonomía	Bueno para medir resultados a los niveles taxonómicos del conocimiento, comprensión y análisis. Inadecuado para medir síntesis y evaluación.	Ineficaz para medir conocimiento. Bueno para medir comprensión, aplicación y análisis. Optimo para medir síntesis y evaluación.
Muestreo del contenido	El uso de un gran número de ítemes permite cubrir una gran variedad de temas, lo que hace posible un muestreo representativo del contenido.	El uso de un número de ítemes relativamente pequeño impide cubrir muchos temas, lo cual obstaculiza la representatividad de la muestra del contenido.
Preparación de ítemes	La preparación de buenos ítemes es difícil y absorbe mucho tiempo.	La preparación de buenos ítemes es difícil, aunque más fácil que la de ítemes de tipo objetivo.
Calificación	Objetivo, simple y muy confiable.	Subjetivo, difícil y menos confiable.
Factores que distorsionan las calificaciones de los estudiantes	Las calificaciones del aprovechamiento de los estudiantes están propensas a ser distorsionadas por la habilidad de lectura y por las respuestas dadas al tanteo.	Las calificaciones de aprovechamiento de los estudiantes están propensas a ser distorsionadas por la habilidad para expresarse en forma escrita y aparentar que se sabe.
Efecto probable sobre el aprendizaje	Alientan a los estudiantes a recordar, interpretar y analizar las ideas de otros.	Alientan a los estudiantes a organizar integrar y expresar sus propias ideas.

C. Rendimiento

Dada la estrecha vinculación que existe entre la educación y la psicología, se observa que la tendencia de la medición va encaminada hacia dos aspectos específicos del individuo:

1. Las capacidades
2. La personalidad

Thorndike y Hagen (1980:33), han dividido a la medición de las capacidades en dos grandes áreas:

- a. Prueba de aptitud: tiene como finalidad medir lo que la persona podría aprender o hacer.
- b. Prueba de aprovechamiento: mide lo que la persona ha aprendido a hacer después de haber recibido determinada clase de enseñanza.

Reafirmando lo anteriormente expuesto:

- 1) Una prueba de aptitud es aquella que se orienta a predecir el éxito en ciertas ocupaciones o aprendizajes.
- 2) Una prueba de aprovechamiento, es una serie de elementos o ítemes, que muestran el conocimiento o dominio que uno o más sujetos posee, con respecto a una o varias unidades de un curso o programa y en un determinado nivel escolar.

Sobre el tema de la medición del aprovechamiento se ha escrito mucho

y así la aplicación de escalas, listas de verificación, cuestionarios y pruebas; son, teóricamente, muy empleados por los docentes como instrumentos de medición del rendimiento.

En la práctica sin embargo, en la mayoría de los casos, es la prueba estructurada el instrumento único para la valoración del aprendizaje.

Si el interés fundamental es obtener información respecto al estado general de un alumno o de un grupo de alumnos, en una destreza o un área de estudios, se elaborará una prueba de exploración.

Si el propósito es diagnosticar las habilidades y las deficiencias específicas del alumno, respecto a una destreza o área de estudio, una prueba de diagnóstico será la indicada.

Es necesario considerar que una prueba solamente proporciona una estimación parcial de los objetivos programados, ya que algunos de ellos no es posible evaluarlos mediante las pruebas existentes.

El presente trabajo plantea la comparación entre parejas de puntuaciones, provenientes de la prueba estructurada; selección múltiple, y de la prueba no estructurada: ensayo. Lamentablemente la estimación de la diferencia entre dos pruebas, en aspectos de confiabilidad, es menor que la confiabilidad de las pruebas tomadas por separado; en ello influyen dos factores:

1. Los errores de medición en ambas pruebas separadas afectan la puntuación de diferencia y

2. Lo común en ambas mediciones se anula en la puntuación de diferencia.

Considerando el rendimiento en matemática, Bloom, Hastings y Madaus (1975:223) señalan:

"La primera preocupación de la evaluación en el aprendizaje de la matemática, ha sido y continuará siendo, los resultados o logros cognoscitivos. Los resultados afectivos sirven de sostén y son importantes, pero son aun secundarios o a lo sumo de igual importancia que el rendimiento.

El rendimiento en matemática posee muchas facetas. Esto significa que no es una estrategia unitaria y que, por consiguiente, es necesaria una estrategia que asegure la consideración de muchas medidas diferentes de rendimiento en matemática. Una estrategia que se ha empleado hasta cierto punto en una variedad de contextos es estratificar los resultados de la instrucción de la matemática de dos maneras: En primer lugar, por tipos de contenido de matemática, y, en segundo lugar, por niveles de conducta".

¿Cuál es el papel de la aptitud del alumno en el rendimiento escolar?

Entre los alumnos existen diferencias individuales y éstas se ponen de manifiesto en las mediciones de aptitud académica y en las de cualquier otro tipo de medición. Es probable que un alumno comprendido en un determinado percentil de aptitud, tenga un aprovechamiento o rendimiento en una posición diferente en dicha escala. Se considera subrendimiento o subaprovechamiento a toda discrepancia entre rendimiento y aptitud, cuando ésta última está por encima del primero. Esto significa que entre las medidas de aptitud y rendimiento no se encuentra una correlación perfecta. Se puede observar que los alumnos que ocupan posiciones altas respecto a sus medidas de aptitud, tendrán un buen rendimiento, pero no necesariamente

obtendrán en aprovechamiento las mayores puntuaciones. Se puede concluir diciendo que la correlación entre aptitud y rendimiento es positiva y estadísticamente significativa, pero no perfecta.

D. Investigaciones similares

Los primeros estudios sobre la confiabilidad de la prueba de ensayo fueron hechos por Starch y Elliot, según Karmel (1975:434):

"En sus investigaciones sobre las matemáticas, enviaron un test en geometría plana a 138 profesores de esta asignatura para que la calificaran, estimando que las matemáticas tendrían mayor consistencia en las calificaciones. Sus resultados mostraron una amplitud de calificaciones entre 28 y 95".

Otra investigación similar fue realizada por Félix Amado De León Reyes, para optar a la Maestría en Medición, Evaluación e Investigación Educativas en la Universidad del Valle de Guatemala en 1977, bajo el título: "Formación docente del profesor de matemática del tercer grado del ciclo básico y los criterios de calificación en una prueba de ensayo".

Se utilizó el método de investigación ex-post-facto: para un estudio de campo encaminado a describir relaciones e interacciones entre las variables sociológicas y educacionales, realizada entre los profesores de matemática de tercer año del ciclo básico de las escuelas públicas y privadas de la ciudad de Guatemala.

El instrumento que sirvió de base a esa investigación fue una prueba de ensayo, diseñada por el autor con base en el programa del tercer año del ciclo básico del Instituto Experimental "Dr. José Matos Pacheco",

la cual aplicó a un grupo de cinco alumnos: A, B, C, D, y E.

Población: Profesores de matemáticas del tercer año del ciclo básico que laboran en institutos de enseñanza básica en el área urbana correspondiente a los sectores público y privado.

Se enviaron copias de las pruebas aplicadas, a una muestra estratificada de 61 profesores, en las calificaciones que remitieron los profesores de los 5 alumnos considerados en esa investigación, se observó lo siguiente: A: rango de 66; B: rango de 56; C: rango de 49; D: rango de 37; y E: rango de 58.

Según Jiménez (1979:36), De León, en una de sus conclusiones dice:

"Existe gran variabilidad entre las calificaciones asignadas por los profesores de matemática del ciclo básico a las pruebas de ensayo".

Lafourcade (1973:77-78) acerca de este tipo de prueba dice:

"Una prueba de ensayo permite al alumno elaborar su propio marco de referencia y seguir, en la solución de una cuestión su propio hilo de pensamiento o su especial manera de resolverla, tal vez no llegue a la solución que se le predetermina.

Las investigaciones respecto a este planteo arrojan algo de luz sobre esta aserción. Ebel (1965:86) apoyándose en un estudio de Cook, expresa que la producción de una respuesta no es necesariamente una tarea más compleja o difícil, o algo más significativo del rendimiento, que la elección de la mejor respuesta entre varias opciones. En su investigación Cook presentó a 152 alumnos del primer año del College dos pruebas de 60 ítems cada una, sobre conocimientos de asuntos contemporáneos. Ambas se referían a las mismas cuestiones, pero en una de ellas el alumno debía pensar y redactar sus respuestas, en cambio en la otra, sólo debía elegir la mejor entre varias opciones. La correlación entre las dos pruebas fue 0.97, lo cual significa que ambas

midieron idénticos aspectos del aprendizaje".

En la Universidad del Valle de Guatemala, Jiménez (1979) realizó un trabajo de investigación a nivel medio titulado: "Comparación de los resultados obtenidos al aplicar pruebas de ensayo y selección a un mismo contenido matemático", él utilizó el método de investigación ex-post-facto, con una población de 56 alumnos del quinto curso de la Escuela Experimental de Aplicación "Mirón Muñoz", con los siguientes instrumentos:

1. Tabla de especificaciones de doble entrada.
2. Prueba estructurada compuesta de 40 ítems de selección múltiple.
3. Prueba de ensayo de 12 ítems.

En relación a la conclusión final del estudio realizado por Jiménez (1979:57)

"No se encuentra diferencia estadísticamente significativa al medir el rendimiento con una "prueba de ensayo" o con una "prueba objetiva".

Los siguientes factores pudieron haber influido en dicha conclusión:

1. La aplicación de las pruebas a fines del curso lectivo y el hecho de que las calificaciones no fueron parte de la evaluación del curso, pudo haber disminuido la motivación del alumno para desarrollar o mostrar su máximo rendimiento.
2. La probable subjetividad de evaluación dado que las pruebas fueron evaluadas únicamente por el autor.

3. La muestra definitiva incluyó a 15 sujetos que participaron en la prueba piloto.

En el estudio realizado por Terrón (1978) bajo el título: "Comentarios del maestro en pruebas objetivas: Sus efectos en el rendimiento".

Terrón utilizó el método de investigación ex-post-facto, siendo su población: 78 alumnos del cuarto grado de tres escuelas primarias de recuperación de la ciudad de Toluca, México. Su instrumento fue una prueba de aprovechamiento en el área de matemática; y usó un diseño experimental que incluyó a dos grupos: experimental y control con Pre y Post-test. Además, aplicó un análisis de covarianza como técnica estadística.

Con base en los resultados que obtuvo aceptó la hipótesis nula:

"El rendimiento del grupo que recibe comentarios del maestro en sus pruebas objetivas (experimental) es el mismo que el del grupo que no recibió comentarios (control), a un nivel de $p < 0.05$ ". Terrón (1978:25)

En relación a su investigación el mismo autor se hizo el siguiente cuestionamiento:

- a. Si el enfoque adoptado para el estudio fue válido; y
- b. Si en el estudio existieron factores que atentaron contra la validez del mismo.

El autor del trabajo optó por la segunda alternativa en base al análisis bibliográfico sobre la motivación y el aprendizaje, planteando que:

- 1) Es posible que la variable experimental (comentarios), no haya producido el efecto deseado, por haberse aplicado una sola vez en el experimento.
- 2) El uso de pruebas no estandarizadas como instrumentos de medición, a pesar de que su confiabilidad, está en los límites de aceptación para pruebas de rendimiento en el aula.
- 3) El experimento fue realizado a fines del ciclo especial.
- 4) Posible interacción del grupo experimental y control.
- 5) Falta de experiencia de los alumnos en el manejo de pruebas de selección múltiple con 4 distractores.

E. Comentario sobre la revisión bibliográfica

De la revisión de la literatura se puede comentar lo siguiente:

1. Existe gran variabilidad entre las calificaciones asignadas por los profesores de matemática a una prueba de ensayo, aspecto que reduce la confiabilidad del instrumento.
2. La prueba estructurada permite obtener una muestra más representativa de la estimación del logro de los objetivos previstos, en relación con la prueba no estructurada. Sin embargo, algunos objetivos no podrán ser estimados mediante la aplicación de ninguno de estos dos tipos de pruebas.
3. No existe un consenso bien definido entre los autores respecto a la clasificación de los ítems, en estructurados y no estructurados.

Además, sus ventajas y desventajas son relativas, dependiendo de los objetivos específicos a evaluar.

4. La correlación entre aptitud y rendimiento es positiva; pero no perfecta.



III. METODOLOGIA

Los aspectos metodológicos utilizados en el presente trabajo se exponen a continuación:

A. Problema

En esta investigación el problema central es dar respuesta a la siguiente pregunta:

¿Existe diferencia en el rendimiento del alumno en matemática al ser medido por una prueba de base estructurada: selección múltiple y una prueba de base no estructurada: Ensayo?

B. Hipótesis

Para las hipótesis que aparecen a continuación se fijó un nivel de significado estadístico de $p_{\alpha} = 0.05$.

H_0 : No existen diferencias estadísticamente significativas entre el rendimiento obtenido por los alumnos al aplicarles pruebas de selección múltiple y pruebas de ensayo, para evaluar los mismos objetivos de una unidad del programa de matemática (MM101).

H_1 : Existen diferencias estadísticamente significativas entre el rendimiento obtenido por los alumnos al aplicarles pruebas de selección múltiple y pruebas de ensayo, para evaluar los mismos objetivos de una unidad del programa de matemática (MM101).

C. Variables

En los párrafos siguientes se enuncian el nombre, la notación y la

definición operacional de las variables.

1. Variable independiente experimental. Es el tipo de prueba que se le aplica a los alumnos para medir su rendimiento en matemática.

Esta variable (X) se aplicó a dos niveles:

- a. X_1 : Primera prueba, de selección múltiple, compuesta de 30 ítemes de cuatro alternativas cada uno, aplicada a nivel del curso de matemática general (MM101) de la Escuela Superior del Profesorado "Francisco Morazán" -ESP- y MM110 de la Universidad Nacional Autónoma de Honduras, -UNAH-, con un tiempo de aplicación de dos horas.
- b. X_2 : Segunda prueba, de ensayo, compuesta de 17 ítemes, aplicada a nivel del curso de MM101 y MM110 con un tiempo de aplicación de dos horas.

2. Variables dependientes. Son los resultados obtenidos de la aplicación de las pruebas

- a. Y_1 : Número de respuestas correctas de los estudiantes en la prueba de selección múltiple.
- b. Y_2 : Número de respuestas correctas de los estudiantes en la prueba de ensayo.
- c. Y_3 : El puntaje, indicado en forma porcentual, obtenido por cada alumno, en la prueba de selección múltiple.

- d. Y_4 : El punteo, indicado en forma porcentual, obtenido por cada alumno en la prueba de ensayo.

D. Diseño experimental

Esta investigación es de tipo "experimento de campo", que Kerlinger (1975:283) lo define así:

"El experimento de campo es un estudio de investigación en una situación real, donde una o más variables independientes son manipuladas por el experimentador bajo condiciones controladas con el máximo cuidado que permita la situación".

El paradigma conceptual del diseño experimental de esta investigación es el siguiente:

a	X_1 — Y	Experimental: Selección múltiple (dos secciones)
	X_2 — Y	Control: Ensayo (una sección)

El cual incluye a un grupo experimental y a un grupo control, con selección de sujetos parcialmente azarizada.

E. Instrumentos

1. Prueba de base estructurada formada por ítemes de selección múltiple, (Anexo B)
2. Prueba de base no estructurada formada por unidades de ensayo, (Anexo C). Ambas midieron los mismos objetivos del programa de matemática MM10 de la Universidad Nacional Autónoma de Honduras y MM101 de la Escuela Superior de la ciudad de Tegucigalpa.
3. Tabla de especificaciones respectiva, (Anexo A). Estas prue-

bas fueron construídas por el autor con base en el programa de estudio del curso de matemática MM110 y MM101 del departamento de matemática de UNAH y ESP correspondientes al año 1982.

Ambos instrumentos usados en este estudio son "pruebas de dominio", las cuales Gronlund (1973:21) define así:

"La mayoría de las pruebas de aprovechamiento son pruebas de dominio, con generosos intervalos de tiempo".

Thorndike y Hagen (1980:678) definen las pruebas de dominio así:

"Test que se emplea para determinar si un alumno o unos alumnos dominan la unidad que se les ha enseñado".

En su forma preliminar, la prueba de selección tenía 50 ítemes, con 4 alternativas y la prueba de ensayo de 20 ítemes, como se explica en la página 37.

Las pruebas definitivas constaron de 30 ítemes de selección múltiple y de 17 ítemes de ensayo. La tabla de especificaciones respectiva se presenta en el Anexo A.

4. Confiabilidad. Partiendo del supuesto de que las pruebas objeto de la presente investigación son homogéneas, por medir una unidad de objetivos y presentar en toda su longitud el mismo tipo de ítemes se pudo calcular su consistencia interna mediante la fórmula de Kuder-Richarson 20, propuesta por Magnuson (1979:143). Se aplicó la fórmula y se obtuvo una confiabilidad de 92 por ciento.

5. Validez

a. Validez interna del diseño experimental. Aspectos tales como los de aplicación de la prueba, instrumentación, sesgos resultantes de la selección de sujetos e interacción entre selección y maduración, teóricamente fueron controlados en este diseño por medio del grupo testigo y la azarización parcial.

b. Validez externa del diseño experimental. La azarización permitió controlar aspectos tales como los efectos de sesgo de la selección; de las variables experimentales, y de una posible interacción entre ambas.

La validez de las pruebas de rendimiento se fundamenta en los objetivos y contenidos programados. Según Karmel (1974:435):

"Los test de aprovechamiento se validan fundándose en si abarcan o no el campo de estudio enseñado y las metas de dicho campo".

Thorndike y Hagen (1980: 678) definen un test de aprovechamiento de la manera siguiente:

"Un test de capacidad que tiene como objeto estimar lo que un individuo ha aprendido a hacer a consecuencia de un entrenamiento o de una experiencia, planeados y previos, que a menudo se proporcionan en la escuela".

6. Evaluación de las pruebas. La prueba de selección múltiple fue calificada asignándole un puntaje de uno en el caso que la respuesta del sujeto coincidiera con la clave correspondiente y cero en el caso contrario. Por lo tanto, el puntaje total del sujeto en la prueba

fue igual al número de respuestas correctas y el valor máximo posible fue de 30. Dicho valor se convirtió en porcentaje respecto al total de puntos de la prueba.

La prueba de ensayo se calificó de 0 a 6, asignando 2 puntos a cada uno de los pasos indicados en las instrucciones de la misma, obteniéndose un porcentaje con relación al total de puntos alcanzados en la prueba.

Ambas pruebas fueron calificadas por el autor del presente trabajo y por los profesores encargados de las secciones seleccionadas. Para la prueba de selección múltiple se usó una clave y para la de ensayo se cifraron a las instrucciones de la misma.

F. Población y muestra

1. La población objeto de este estudio la constituyeron los alumnos de primer ingreso a la Escuela Superior "Francisco Morazán" que cursaron la asignatura MM101, durante el año 1982.

2. La muestra la constituyeron cuatro secciones de 48 alumnos cada una, seleccionadas al azar por el método aleatorio simple, de un total de 8 secciones. Las cuatro secciones: A, B, C, y D se distribuyeron al azar de la siguiente forma:

- a. Prueba piloto (una sección)
- b. Grupo control (otra sección)
- c. Grupo experimental (las otras dos secciones)

3. El estudio piloto se planificó con el fin de probar los instrumentos a emplear en la experimentación definitiva. Para llevarlo a cabo se seleccionó una sección al azar (40 alumnos), con características muy similares a las de la población que se investigó en el estudio y se les aplicó el instrumento inicial.

En base a esta aplicación previa, se realizó el análisis de los 50 ítemes de la prueba original de selección múltiple y los de la prueba de ensayo de 20 ítemes. El instrumento final de selección múltiple se integró con los 30 ítemes de selección que presentaron mayores coeficientes estadísticos de facilidad y de discriminación. En la misma forma se elaboró la prueba de ensayo con 17 ítemes.

4. El estudio definitivo para la evaluación de los objetivos de la unidad: "Conjunto de los números reales", del programa de matemática I (MA1101), se realizó mediante la aplicación de las dos pruebas elaboradas, a la muestra representativa de la población seleccionada.

G. Prueba estadística

La técnica estadística que se utilizó para someter a prueba las hipótesis en el presente trabajo, fue la comparación de medias independientes, por medio de una "t" de Student.

H. Procedimiento

El procedimiento seguido para la recolección de los datos fue:

1. Análisis de la población disponible para seleccionar la que reunía las condiciones establecidas para realizar el estudio; es decir,

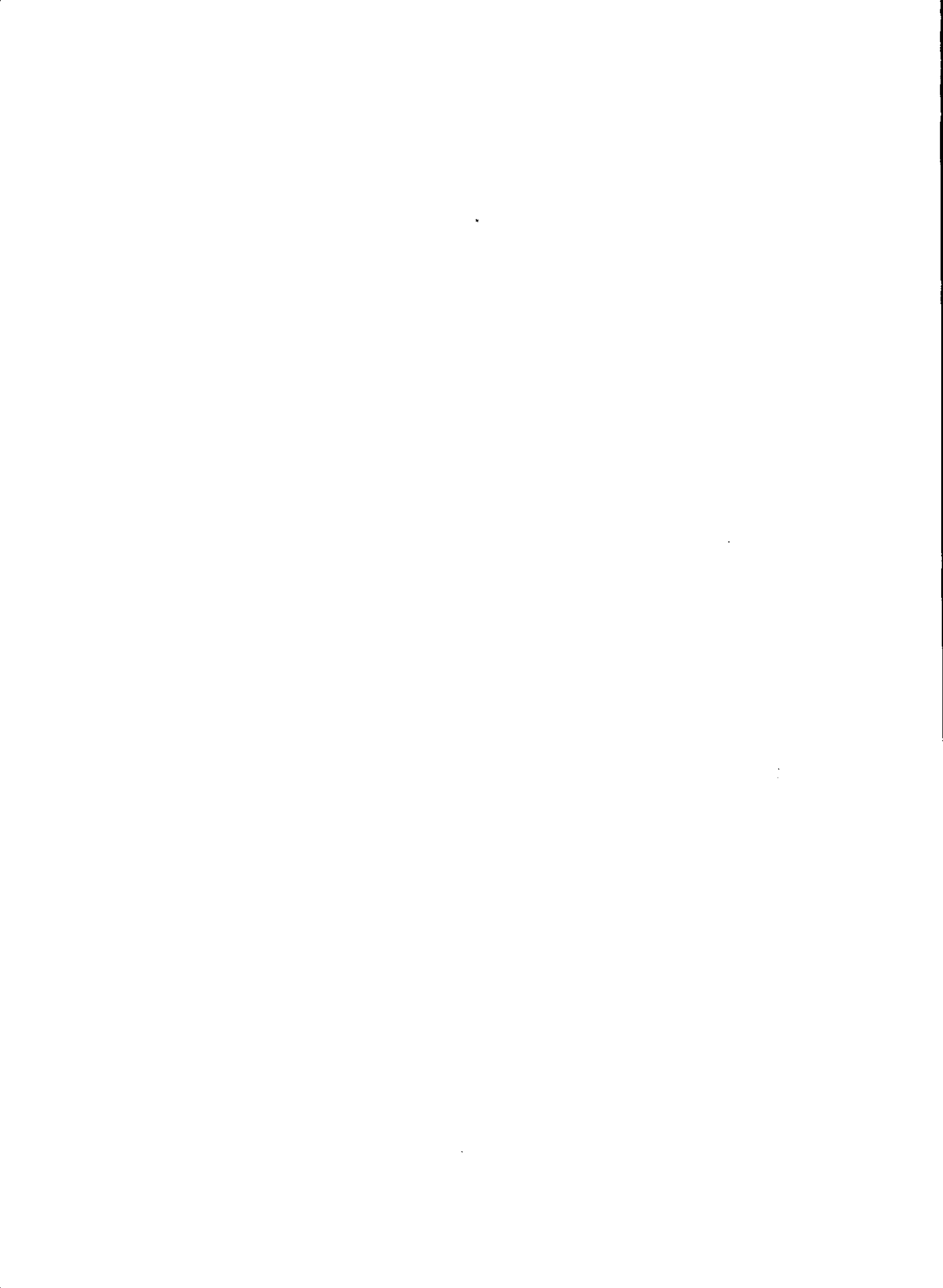
que se estableció la cobertura de la población a investigar.

2. Cálculo de la muestra representativa de esa población.
3. Construcción de dos pruebas experimentales basadas en una unidad del programa de MM101 del departamento de matemática de la Escuela Superior "Francisco Morazán".
4. Elaboración de los ítemes de estas pruebas por el autor de esta investigación de acuerdo a:
 - a. Los objetivos educativos presentados por Gronlund (1973:600-604).
 - b. El programa de MM101 del departamento de matemática de la Escuela Superior.
 - c. Materiales de apoyo.
 - d. La información suministrada por los profesores de los cursos respectivos.
5. Aplicación de la prueba piloto a una sección de 40 alumnos seleccionados al azar; calificación de las pruebas y análisis de los ítemes de acuerdo a su grado de dificultad y de discriminación.
6. Selección de los ítemes de las pruebas definitivas, en base a los resultados de su análisis respectivo.
7. Asignación al azar de los tratamientos a las tres secciones, y aplicación de las pruebas definitivas de la siguiente manera:

- a. Una sección de grupo control; a las filas pares se les dió la prueba de selección múltiple y a las impares la prueba de ensayo.
- b. Dos secciones de grupo experimental. La primera sección de éstas, se dividió en filas pares e impares. A las pares se les asignó la prueba de ensayo y a las impares la prueba de selección y viceversa en la segunda aplicación de las pruebas.

La segunda sección, se organizó así: a las filas pares se les asignó la prueba de selección y a las impares la prueba de ensayo y viceversa en la segunda aplicación.

8. Calificación, análisis de ítemes y cálculo de los coeficientes estadísticos de las dos formas de las pruebas aplicadas.
9. Tabulación de los resultados en cuadros que permitieron una visión global de los mismos.
10. Análisis e interpretación de los resultados.



IV. RESULTADOS

En el presente capítulo se dan las estadísticas descriptivas e inferenciales resultantes del análisis de los datos correspondientes a la investigación.

A. Cuadros informativos

En el cuadro 4.1 se presenta los resultados obtenidos por los estudiantes en la prueba de selección múltiple y en la de ensayo.

En el cuadro 4.1 se puede observar que los estudiantes con mayores puntuaciones en la prueba de selección múltiple, no superaron los puntajes respectivos obtenidos en la prueba de ensayo salvo en un caso; sin embargo, el porcentaje de aprobación de los estudiantes en la prueba de selección (64%) superó al correspondiente de la prueba de ensayo (41%).

En el cuadro 4.2 se presenta la confiabilidad de las pruebas de selección múltiple y de ensayo.

El valor de confiabilidad de la prueba de selección múltiple tiene dos alternativas en su cálculo:

1. para punteos burdos con $K-R_{20}$ y $K-R_{21}$; y
2. porcentajes con $K-R_{20}$ y $K-R_{21}$; donde, $K-R_{20}$ y $K-R_{21}$ identifican la "fórmula 20" y "fórmula 21" de Kuder-Richardson para la confiabilidad de un test con n ítemes; según (Magnuson: 144-145).

Cuadro 4.1

Punteos burdos y porcentajes obtenidos por los estudiantes en la prueba de selección múltiple; y porcentajes en la prueba de ensayo (n = 22)

Sujeto	Tipos de prueba		
	Selección múltiple		Ensayo
	Punteos burdos	Porcentajes	Porcentajes
1	19	63	74
2	20	67	42
3	24	80	75
4	20	67	53
5	13	43	29
6	08	27	15
7	22	73	94
8	19	63	65
9	18	60	52
10	14	47	35
11	19	63	51
12	21	70	95
13	19	63	44
14	12	40	43
15	14	47	47
16	14	47	30
17	17	57	67
18	24	80	98
19	19	63	69
20	19	63	92
21	19	63	24
22	17	57	38

Al establecer comparaciones se observa que los resultados, en términos de porcentajes, elevan el coeficiente de confiabilidad de la prueba de selección; resultados que sobrepasan los límites de confiabilidad establecidos para un test elaborado por el maestro de acuerdo al número de ítemes; según (Lafourcade:183). Los coeficientes de confiabilidad para puntajes burdos y su equivalente en porcentajes, obtenidos mediante $K-R_{21}$ también fueron verificados en el Centro de Computación de la Universidad del Valle de Guatemala.

El coeficiente de confiabilidad para la prueba de ensayo se calculó con la fórmula $K-R_{20}$, que es un caso especial del coeficiente alfa (∞) de Cronbach, usado para estimar la confiabilidad de un test o estimaciones donde los ítemes dan puntajes pesados (Magnuson:145), como en el presente caso.

El coeficiente de confiabilidad aceptable para una prueba elaborada por el maestro con 20 ítemes es de 0.50 y con 40 ítemes es 0.67, según (Lafourcade:183). De acuerdo a los resultados del cuadro 4.2 para la prueba de selección múltiple elaborada por el autor con 30 ítemes, en ambos casos tiene una confiabilidad aceptable, para permitir el análisis comparativo con el rendimiento medido con la prueba de ensayo.

Cuadro 4.2
Confiabilidad de las pruebas

Tipo de prueba	Confiabilidad									
Ensayo	0.89									
Selección	<table style="margin: auto; border: none;"> <thead> <tr> <th style="text-align: left;"><u>Punteos burdos</u></th> <th style="text-align: center;">← KR₂₀ →</th> <th style="text-align: right;"><u>P. en porcentaje</u></th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">rtt = 0.77</td> <td></td> <td style="text-align: center;">rtt = 0.92</td> </tr> <tr> <td style="text-align: center;">rtt = 0.68</td> <td style="text-align: center;">← KR₂₁ →</td> <td style="text-align: center;">rtt = 0.86</td> </tr> </tbody> </table>	<u>Punteos burdos</u>	← KR ₂₀ →	<u>P. en porcentaje</u>	rtt = 0.77		rtt = 0.92	rtt = 0.68	← KR ₂₁ →	rtt = 0.86
<u>Punteos burdos</u>	← KR ₂₀ →	<u>P. en porcentaje</u>								
rtt = 0.77		rtt = 0.92								
rtt = 0.68	← KR ₂₁ →	rtt = 0.86								

B. Prueba estadística

La prueba "t de student" varía de fórmula dependiendo de la homogeneidad de la varianza, del tamaño y de la correlación e independencia de las muestras.

En la presente investigación se utilizó la fórmula II B de: (Downie y Heath:200) para el contraste de diferencias entre medias de los resultados de la prueba de selección múltiple y la prueba de ensayo.

En el cálculo de la "t de student" requerida se siguió el procedimiento utilizado por: (Downie y Heath:191-192). El resumen se presenta en el cuadro 4.3.

La "t" crítica con 21 grados de libertad para $p_c = 0.05$ es de 2.080

Cuadro 4.3

Resumen de la prueba "t de Student" para contrastar la significación de la diferencia entre dos medias

Prueba	N	\bar{X}	S	S ²	$t_{\bar{x}_A - \bar{x}_B}$
Selección	22	59.23	12.95	167.70	0.9
Ensayo	22	56.00	24.49	599.76	

C. Interpretación de resultados

De acuerdo al cuadro 4.3 con $p_{\alpha} = 0.05$, el resultado de $t_{\bar{x}_A - \bar{x}_B}$ es menor que el de t crítico y por consiguiente, no se rechaza la hipótesis nula.



V. DISCUSION DE LOS RESULTADOS

En esta parte, se procede a formular las conclusiones sobre el análisis estadístico presentado en el capítulo anterior; se resume las limitaciones encontradas para llevar a cabo el estudio, y se ofrece algunas recomendaciones en base a los resultados encontrados en la investigación.

A. Conclusiones

En este trabajo se comparó los resultados obtenidos al aplicar pruebas de ensayo y pruebas de selección para evaluar los mismos objetivos de un programa de matemáticas de nivel superior. Se seleccionó la unidad: El conjunto de los números reales, del programa de matemática I (Mat 101) que cursan los alumnos de primer ingreso a la Escuela Superior del Profesorado: "Francisco Morazán", estableciéndose de acuerdo al cuadro 4.3 la siguiente conclusión: Probablemente con un nivel de significación de error alfa igual a 5 por ciento ($\alpha = 0.05$), la hipótesis nula, H_0 : "No existen diferencias estadísticamente significativas entre el rendimiento obtenido por los alumnos al aplicarles pruebas de selección múltiple y pruebas de ensayo, para evaluar los mismos objetivos del programa de Matemática I" por lo tanto, no se rechaza.

Ante la pregunta ¿Cuál de los dos tipos de prueba es más eficiente? Será necesario considerar el concepto de validéz, en el sentido de que la validéz mide la conducta específica en el objetivo sometido a evaluación; por consiguiente, será necesario conocer previamente qué conducta se espera medir. Ningún instrumento de medición es una receta única ya

que cada tema a enseñar ofrece al docente la oportunidad de desarrollar su iniciativa y creatividad, respecto a la situación y a las condiciones particulares del grupo de estudiantes a su cargo. Así, en conclusión es el docente el que tiene que tomar la decisión acerca del instrumento de medición más adecuado a utilizar, atendiente los objetivos previstos. Sobre esta situación se hizo mención en la fundamentación teórica en las páginas-20-21.

B. Limitaciones

1. De acuerdo al plan de trabajo elaborado para la realización del estudio, en el mes de agosto se aplicarían las pruebas de ensayo y de selección a la muestra representativa de la población de los alumnos que ingresan a la Escuela Superior y cursan Matemática I, en Tegucigalpa. Por motivos de fuerza mayor, fuera del control del autor del presente trabajo dicha planificación se realizó sólo parcialmente en la fecha prevista, y hubo necesidad de hacer ajustes para completarla de acuerdo al año lectivo del Programa de Maestría de la Universidad del Valle de Guatemala. En el plan inicial del estudio se había contemplado incluir en la investigación a los estudiantes de la Universidad Nacional Autónoma de Honduras -UNAH- que cursan la materia MM110, pero no se pudo recurrir a dicha población debido a los problemas administrativos que hicieron coincidir el mes de agosto del presente año con el período intersemestral, en el cual no hubo actividades de aula. Por lo tanto los resultados del trabajo, sólo podrán en este caso, ser generalizados a la población de la cual se extrajo la muestra o a poblaciones que tengan las mismas características.

2. La muestra representativa se vio reducida al 20.43% ($n = 22$), debido a que algunas pruebas no fueron desarrolladas en su totalidad y a otros factores fuera del control del autor.

C. Recomendaciones

1. Tratar que en futuros estudios similares la prueba de selección múltiple y la prueba de ensayo formen parte de los exámenes regulares de aprovechamiento, en la evaluación de los participantes.
2. Siempre que se utilice las pruebas de ensayo en el proceso evaluativo se formen grupos de profesores jueces, según la especialidad para evaluarlas, con el propósito de reducir la subjetividad que afectaría las variables en estudio. Además, deberá elaborarse una clave a fin de que los profesores jueces ignoren la identidad de los estudiantes.
3. Extender este tipo de estudios, a las demás áreas de formación de la Escuela Superior y de la Universidad Nacional.
4. Informar detalladamente al personal de la institución de la cual se obtendrá la muestra, sobre los objetivos y la importancia de la investigación en general, a fin de lograr actitudes positivas y la máxima cooperación.
5. Responsabilizar a los coordinadores de área, de archivar las pruebas aplicadas anualmente, con una proyección futura de seleccionar los ítemes más adecuados en cuanto a índices de discriminación y de dificultad, y poder así elaborar instrumentos con mayor validez y confiabilidad.



BIBLIOGRAFIA

- Adams, G. S. Medición y Evaluación en Educación, Psicología y Guidance.
1970 Barcelona, Herder.
- Bloom, Benjamín; Hastings, Thomas y Madaus George. Evaluación del Aprendizaje.
1975 Buenos Aires, Troquel.
- Cronbach, J. L. Fundamentos de la exploración psicológica. Madrid,
1972 Editorial Biblioteca Nueva.
- De León, Félix A. Formación docente del profesor de matemática del tercer grado del ciclo básico y los criterios de calificación en una prueba de ensayo. Guatemala, Universidad del Valle de Guatemala.
- Downie, N. M. y R. W. Heath. Métodos Estadísticos Aplicados. México,
1973 Harla.
- Goring, A. Paul. Manual de Mediciones y Evaluación. Buenos Aires,
1971 Kapelusz.
- Gronlund, Norman. Elaboración de tests de aprovechamiento. México,
1974 Trillas.
- Jiménez Villalobos, Francisco. Comparación de los resultados obtenidos al aplicar pruebas de ensayo y selección a un mismo contenido matemático. Guatemala, Universidad del Valle de Guatemala.
- Karmel, Louis. Medición y Evaluación Escolar. México, Trillas.
1974
- Kerlinger, Fred. Investigación del comportamiento. México, Interamericana.
1975
- Lafourcade, Pedro. Evaluación de los aprendizajes. Buenos Aires, Kapelusz.
1973
- Magnuson, David. Teoría de los Tests. México, Biblioteca Técnica de Psicología, Trillas.
1979
- Morales Muñoz y Valvuela. Manual sobre las pruebas de rendimiento escolar. Caracas, Ediciones Co. Bo.
1978
- National Council of Teachers of Mathematics. Evaluation in Mathematics.
1965 Washington.
- Nilo, Sergio U. "Temas de Evaluación". Revista, Educación Hoy. Año III (17): 5-6 Bogotá, Colombia

Terrón Estrada, Jorge. Comentarios del maestro en pruebas objetivas: sus efectos en el rendimiento. Guatemala, Universidad del Valle de Guatemala.

Thorndike, Robert y Elizabeth Hagen. Tests y Técnicas de Medición en Psicología y Educación. México, Trillas.

ANEXO A

Tabla de especificaciones

Objetivo	Prueba	Conocimiento 23%	Comprensión 47%	Aplicación 30%	Análisis
1. Definir el conjunto de los números reales como $R=QUI$	Estructurada No estructurada	1			
2. Definir la relación mayor que, menor que, mayor o igual que y menor o igual que; en R.	Estructurada No estructurada	28	2, 3 7		
3. Escribir un subconjunto de R. en notación de intervalos	Estructurada No estructurada		4, 6 6B	5	
4. Realizar las operaciones de Unión, Intersección, diferencia y complemento de intervalos reales y graficarlas en la recta numérica	Estructurada No estructurada		6A, 6C		
5. Definir las operaciones de adición, sustracción, multiplicación, división y potenciación en R.	Estructurada No estructurada	9, 21, 23 10, 12A	14, 24 2	11, 26, 27 3A, 14, 17	
6. Aplicar las propiedades de adición, sustracción, multiplicación, división en R.	Estructurada	25	12, 13, 17, 19, 28	15, 30	
7. Aplicar las propiedades de las potencias con exponente racional	No estructurada Estructurada		1, 3B, 5 20	15 7, 10, 16	16
8. Resolver problemas que utilicen notación científica aplicando propiedades de las potencias	No estructurada Estructurada	8	8, 2B		
9. Simplificar raíces	Estructurada No estructurada		22 13		
10. Calcular el lado de un triángulo rectángulo conociendo los otros	Estructurada No estructurada		18 4		



ANEXO B

PRUEBA ESTRUCTURADA

Nombre del alumno: _____

No. de Carnet: _____ No. Asignado: _____

Fecha: _____

INSTRUCCIONES:

A continuación encontrará **30** ítemes de selección múltiple, de 4 opciones cada una, seleccione la respuesta más acertada, encerrando en un círculo la letra que acompaña la opción correspondiente.

Valor de cada ítem: 1%
Tiempo estimado: 2 horas

- 1.- El conjunto de los números reales está formado por:

a. $Q \cap I$	c. $Q \cup Z$
b. $Q \cup I$	d. $Q \cap Z$

- 2.- Si $a > b$, $a > d$ y $b > c$ ¿Cuál de las siguientes desigualdades es correcta?

a. $a > c$	c. $c > a$
b. $c > d$	d. $d > c$

- 3.- ¿Cuál de los siguientes conjuntos numéricos está ordenado en forma ascendente?

a. $\left\{ \frac{5}{8}, \frac{10}{11}, \frac{2}{7} \right\}$	c. $\left\{ \frac{2}{7}, \frac{5}{8}, \frac{10}{11} \right\}$
b. $\left\{ \frac{10}{11}, \frac{2}{7}, \frac{5}{8} \right\}$	d. $\left\{ \frac{2}{7}, \frac{10}{11}, \frac{5}{8} \right\}$

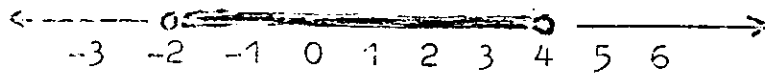
- 4.- Utilizando la notación de intervalos, el conjunto $W = \left\{ x/x \in R, -2 < x \leq 5 \right\}$ se escribe:

a. $] -2,5 [$	c. $[5, -2 [$
b. $[-2,5 [$	d. $] -2,5]$

- 5.- La intersección del conjunto $K = \left\{ x/x \in R, -3 \leq x \leq 3 \right\}$ y el conjunto $H = \left\{ x/x \in R, x \geq 2 \right\}$ es:

a. $] 2,3 [$	c. $[2, + \infty [$
b. $[2,3]$	d. $[-3,3]$

6.- ¿A qué conjunto corresponde la gráfica



- a. $\{x/x \in \mathbb{R}, -2 < x < 4\}$ c. $\{x/x \in \mathbb{R}, -2 < x \leq 4\}$
 b. $\{x/x \in \mathbb{R}, -2 \leq x < 4\}$ d. $\{x/x \in \mathbb{R}, -2 \leq x \leq 4\}$

7.- En la expresión $\frac{r^2 - s^2}{rs} = 0$, r y s son números positivos
 ¿Cuál es el valor de "r" si "s" vale 1?

- a. 1 c. -1
 b. 2 d. 0

8.- El resultado de multiplicar 0.085 por 1,000 es:

- a. 8,500 c. 85
 b. 850 d. 8.5

9.- ¿Cuál es el resultado al efectuar $\frac{2}{\sqrt{3}} \cdot \frac{3}{\sqrt{2}}$?

- a. $\frac{6}{\sqrt{5}}$ c. $\frac{5}{\sqrt{3+\sqrt{2}}}$
 b. $\frac{\sqrt{6}}{6}$ d. $\sqrt{6}$

10.- El valor de la expresión $(X^X)^{(X^X)}$ cuando x=2 es:

- a. 16 c. 256
 b. 64 d. 1024

11.- El resultado de operar $\frac{2}{3} \left(\frac{1}{2} + \frac{1}{4} \right) \div \left(\frac{1}{2} - \frac{1}{5} \right)$ es

- a. $\frac{3}{5}$ c. $\frac{5}{3}$
 b. $\frac{7}{10}$ d. $\frac{6}{7}$

12.- $3(5\sqrt{3} + 4\sqrt{5}) = 15\sqrt{3} + 12\sqrt{5}$ por la propiedad

- a. Asociativa c. Distributiva
 b. Conmutativa d. Ninguna de las anteriores

13.- El inverso multiplicativo de $\frac{2}{\sqrt{3}}$ es

- a. $\frac{3}{\sqrt{3}}$ c. $\frac{3\sqrt{3}}{2}$
 b. $\frac{2\sqrt{3}}{3}$ d. $\frac{3}{2\sqrt{3}}$

23.- Si $\frac{8}{3} = \frac{x}{9}$, x es igual a:

- a. 16
b. 24
c. 27
d) 72

24.- Tres obreros ganan salario diferente.- El primero gana "x" lempiras, el segundo "y" lempiras y el tercero gana "z" lempiras. ¿Cuál es el salario promedio que ganan estos tres obreros?

- a. $\frac{x + y + z}{3}$
b. $x + y + z$
c. $x \cdot y \cdot z$
d. $\frac{x + y + z}{3}$

25.- La expresión $0 \cdot x$ es igual a:

- a. x
b. x^{-1}
c. 0
d. Infinito

26.- Al efectuar $\frac{2 \cdot \frac{2}{5}}{4\sqrt{\frac{1}{4}}}$ obtenemos:

- a. $\frac{2}{5}$
b. $1 \frac{6}{10}$
c. $\frac{8}{15}$
d. $1 \frac{3}{5}$

27.- En una clase hay 50 alumnos.- Si las dos quintas partes de los alumnos son mujeres y la mitad de las mujeres tienen pelo negro ¿Cuántas alumnas tienen pelo negro?

- a. 30
b. 25
c. 20
d. 10

28.- ¿Cuál de los siguientes decimales es el mayor?

- a. 0.01
b. 0.1
c. 0.9
d. 0.09

29.- Si M, N, P pertenecen al conjunto de los números reales $(M+N)P$ es igual a:

- a. $MP+NP$
b. $MP+MN$
c. $MP(M+N)$
d. $MP+NP$

30.- $cd+md$ es equivalente a:

- a. $(c+m)d$
b. $(c+d)(m+d)$
c. $(c+d)m$
d. $(d+m)c$

ANEXO C

PRUEBA NO ESTRUCTURADA

Nombre del alumno: _____

No. de Carnet: _____ No. Asignado: _____

Fecha: _____

INSTRUCCIONES

A continuación encontrará 17 ítemes de ensayo. - Efectúe en las páginas anexas el planteo, desarrollo y respuesta en forma clara y ordenada.

Valor de cada problema: 6%

Tiempo estimado: 2 horas

1.- Escriba el nombre de la propiedad ejemplificada en cada una de las expresiones siguientes:

a. $5 \cdot (3+6) = (3+6) \cdot 5$ _____

b. $2(\sqrt{2} + 3\sqrt{5}) = 2\sqrt{2} + 6\sqrt{5}$ _____

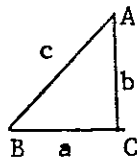
c. $\sqrt{2} + 0 = 0 + \sqrt{2} = \sqrt{2}$ _____

2.- Evaluar la expresión $- [x+(-y)]$ si $x=5$, $y= -3$

3.- Simplificar: a) $\frac{18}{35} + \frac{12}{25}$

b) $\frac{1}{4} \cdot \frac{3}{8\sqrt{2}} \cdot 16\sqrt{32}$

4.- Si



$a=5$ $b=5\sqrt{3}$, entonces $c=?$

5.- a) $2\sqrt{18} - \sqrt{50}$

b) $5\sqrt{3} - 2\sqrt{3} + \sqrt{3}$

6.- Dado $K = \{x/x \in \mathbb{R}, -3 \leq x \leq 3\}$ y $M = \{x/x \in \mathbb{R}, x \geq 2\}$

Determinar mediante la notación de conjuntos (caracterización por comprensión)

A. a. $K \cup M$

b. $K \cap M$

B. Escribir las respuestas a y b del inciso A en notación de intervalos

C. Representar las respuestas a y b del inciso A en la recta numérica

7.- Ordenar el conjunto $\{5/8, 10/11, 2/7\}$ en forma ascendente

8.- Efectuar $(8,75 \times 10^3) + (225 \times 10^{-2})$

- 9.- Determinar el valor de la expresión $(Y^Y)^{(Y^Y)}$ cuando $Y=2$
- 10.- Simplificar $\frac{8\sqrt{18}}{3\sqrt{2}}$
- 11.- Efectuar a) $2^{3/4} \cdot 5^{3/4}$ b) $3^{2/5} \cdot 3^{3/5}$ c) $(\frac{4}{9})^{3/2}$
- 12.- Efectuar a) $136.09 \div 43.9$ b) $\frac{0.04 \times 24}{0.4}$
- 13.- Efectuar aplicando las propiedades de los radicales
- a) $\sqrt{\frac{4}{9} \cdot \frac{25}{36}}$ b) $\sqrt[3]{\frac{1}{8} \cdot \frac{125}{64}}$
- 14.- ¿Cuál es el resultado al operar $\frac{2}{3} (\frac{1}{2} + \frac{1}{4}) \div (\frac{1}{2} - \frac{1}{5})$
- 15.- En un test de matemática un estudiante obtuvo 32 en una escala de 1 a 40 ¿Cuál será su calificación en una escala de 1 a 100?
- 16.- Simplificar la expresión
- $$\frac{(x^{-2} y^4)^3}{(xy)^{-3}}$$
- 17.- En una clase hay 60 alumnos.- Si las $\frac{2}{5}$ partes de los alumnos son mujeres y la mitad de ellas tienen pelo negro. ¿Cuántas alumnas tienen pelo negro?

3/8/82
/mms.



