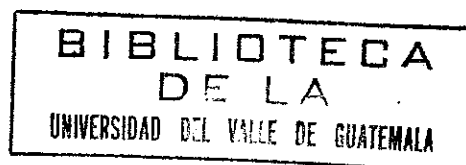


UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Educación



DISEÑO DE UN MODELO PARA ANALIZAR PRUEBAS

OBJETIVAS DE AULA EN FORMA MANUAL

FAUSTINO C. BOTACIO TERREROS

"

TOMÁS CASTELLANOS AVELLANEDA

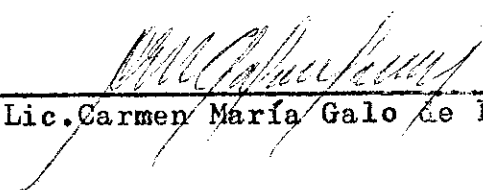
FERNANDO GONZÁLEZ AGUILAR

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


Guatemala

1975

Vo. Bo. de la Coordinadora:

(f) 
Lic. Carmen María Galo de Lara

Vo. Bo. del Asesor:

(f) 
Lic. Tomás José Barrientos González

"Al tomar en la mano un volumen cualquiera... preguntémosnos: ¿Contiene algún razonamiento abstracto con respecto a cantidad o número? No. ¿Contiene algún razonamiento experimental referente a hechos y la existencia? No. Entonces entreguémoslo a las llamas: ;pues no puede contener otra cosa que sofistería e ilusión!"

David Hume (1777)⁺

+ Tomado de Garret, 1971

ÍNDICE

	Páginas
I. INTRODUCCIÓN	1
II. EL PROBLEMA	3
A. Planteamiento	3
B. Antecedentes y justificación	4
C. Objetivos	9
D. Definición operacional de términos	10
III. REVISIÓN DE LA LITERATURA	13
A. Comentario de las obras relevantes	14
IV. MANUAL DE TRABAJO	27
A. Fundamentos teóricos	27
B. Descripción de formatos de trabajo e instrucciones para su manejo	35
C. Presentación de fórmulas estadísticas	43
V. TRABAJO DE CAMPO	47
VI. CONCLUSIONES Y RECOMENDACIONES	49
VII. BIBLIOGRAFÍA	51
APÉNDICES	53
A. Encuesta de evaluación	53
B. Tabla para la determinación del tamaño de n	57
C. Tabla de cuadrados de números y raíces de cero a cincuenta	59

	Páginas
D. Tabla para la determinación de la relación de proporciones $\sqrt{PI/QI}$ para el cálculo de las correlaciones biseriales ítem-test	61
E. Tabla para la determinación de $K/(K-1)$	63
F. Diagrama en secuencia de operaciones para la aplicación del modelo	65
G. Tabla de decisiones e interpretación de resultados	67

I. INTRODUCCIÓN

Interesados en el mejoramiento de la evaluación educativa y, colateralmente, en el de la enseñanza en general, concebimos este manual como un instrumento que sirva al maestro para analizar y evaluar estadísticamente las pruebas objetivas de aula.

Se presenta una base teórica que ayude a comprender los términos utilizados en el manual y también un recuento cronológico de algunas concepciones y de los métodos que han utilizado diferentes autores a través del tiempo. Los cálculos en que se basa el análisis se reducen a llenar dos formatos de trabajo siguiendo instrucciones específicas. Después de haber realizado todos los cálculos aritméticos que requiere el manual, aportará información sobre: grado de facilidad de cada pregunta, confiabilidad de la prueba, poder de separar (discriminar) entre los alumnos con alto rendimiento y los de bajo rendimiento, la posible explicación lógica del mal funcionamiento de las alternativas de la prueba (ambigüedades, errores de redacción, errores técnicos, etc.). Finalmente, se presenta una tabla de decisiones (fundamentada en las técnicas de algunos autores) que ayude a decidir cuándo se debe eliminar un ítem, cuándo mejorarlo, cuándo conservarlo y hasta qué punto la prueba mide lo que intenta medir.

Las ventajas inmediatas de analizar una prueba son: perfeccionar al maestro en la construcción de buenas pruebas, la posibilidad de mejorar la metodología de enseñanza, rehacer algunos temas, brinda retroalimentación en la labor de enseñanza y aprendizaje a estudiantes, a maestros, a la administración escolar, etc.

Este manual va dirigido a todos aquellos educadores conscientes de su labor que pese a la ardua tarea que la enseñanza implica, se interesan por depurar día a día los productos de la interacción indisoluble enseñanza-aprendizaje. Esperamos que sea de utilidad práctica para los educadores y que su uso se extienda a todos los niveles.

II. EL PROBLEMA

El interés por realizar este trabajo surge de la vivencia observada en la educación, en todos los niveles, de una deficiente evaluación de los propios instrumentos de evaluación en el campo de la educación.

A. Planteamiento

Actualmente, el educador cuenta con limitados mecanismos metodológicos que le permitan hacer un balance objetivo de la evaluación del proceso enseñanza-aprendizaje, del cual es él mismo parte activa y preponderante. Al maestro corresponde la responsabilidad de medir el progreso de sus alumnos con respecto a los objetivos del programa de enseñanza. Por esto, debe construir sus propios instrumentos para medir el progreso del alumno en esa dirección.

Si el maestro contara con una guía sencilla para hacer un análisis de la transmisión y recepción de conocimientos, podría actuar sobre aquellos puntos que requieren debida atención, a fin de mejorar los resultados del aprendizaje. La guía debe contener cálculos estadísticos a su alcance que le permitan descubrir fallas en la interacción maestro alumno, errores técnicos en la construcción de los ítemes, etc.

El propósito de este trabajo es brindar al maestro una guía práctica que le sirva para apreciar sus logros y fallas

tanto en los resultados del aprendizaje como en la elaboración de sus ítemes y tests, haciendo acccesibles y manejables los métodos técnicos utilizados para obtener los índices estadísticos adecuados en la toma de decisiones a nivel de aula, bien sea con respecto al rendimiento de sus alumnos o bien con respecto a la calidad técnica de su instrumento de medición o a ambas cosas.

B. Antecedentes y justificación

Entre los múltiples elementos que constituyen el proceso enseñanza-aprendizaje, y como uno de los componentes principales del mismo, se encuentra la serie de procedimientos que las maestros emplean para determinar el grado de aprendizaje logrado en los estudiantes.

Los sistemas de evaluación del aprendizaje, tendrán diferentes características que dependen de la intención que se tenga al ponerlos en práctica. Así por ejemplo, las características de dos sistemas de evaluación diferirán considerablemente si su intención es asignar notas para la promoción o si, por el contrario, se pretende mejorar los métodos de enseñanza, basando las decisiones de cambio en los resultados de la evaluación del aprendizaje. Sin embargo, hay un punto en el cual coincidirán siempre los sistemas de evaluación escolar, independientemente de la utilización que se de a los resultados

obtenidos y es que los instrumentos empleados deberán ser lo suficientemente válidos y confiables para poder tomar decisiones acertadas basadas en los datos obtenidos en la evaluación.

Muchas preguntas pueden hacerse para cuestionar la calidad de los instrumentos de evaluación utilizados a nivel de aula. Como ejemplo, podemos formular las siguientes: ¿Cubre un test los contenidos señalados por los programas de estudio? ¿Lo hace en función de las habilidades y destrezas que los programas de estudios se proponen desarrollar en el estudiante? ¿Se puede estar seguro de que un test mide realmente lo que se propone? ¿Dará un test resultados semejantes si se aplica a grupos con niveles semejantes de habilidad? Contestar a estas y otras preguntas puede resultar imposible o erróneo, si no se fundamentan las respuestas en un análisis que permita cuantificar las características del test.

Una técnica que resulta de gran utilidad, en caso de ser utilizada, es la que podríamos denominar "Análisis de Pruebas". Sin embargo, en nuestros sistemas educativos, formales e informales, es muy pobre el conocimiento que tienen los educadores del análisis de pruebas y, acaso en algunos países, se utiliza exclusivamente a nivel de programas

nacionales cuyos resultados poco inciden en el proceso de la enseñanza, porque se usan con fines específicos de selección o promoción masivas. Estamos conscientes de que este tipo de análisis es más funcional a nivel de aula, ya que el maestro puede calibrar y dosificar la enseñanza en la medida en que lo indiquen los resultados objetivos.

Para hacer un análisis completo de un test y poder elaborar algunas inferencias válidas, se necesita de mucho tiempo y de suficientes conocimientos básicos de estadística que permitan el manejo de conceptos, fórmulas, gráficos, tablas, etc. Este es el motivo por el cual el educador considera este tipo de trabajo, reservado para una computadora electrónica o para un equipo de expertos en la materia.

Por otro lado, creemos que es posible simplificar toda la estructura formal del procedimiento, reduciendo el tedioso trabajo algebraico a sencillos procedimientos aritméticos factibles en poco tiempo y reducido espacio.

Los resultados empleados para evaluar el grado de aprovechamiento de los conocimientos transmitidos en la escuela, forman parte importante de la empresa educativa. Al hacer un análisis completo de un test objetivo, el maestro está en condiciones de determinar si el conocimiento transmitido fue claro y logró su objetivo. De no ser así, se puede preguntar:

¿Qué fue lo que interpretó el estudiante respecto a la forma de expresión de un contenido dado? ¿Por qué algunas preguntas no funcionaron como se esperaba? ¿Qué parte del programa produjo confusión entre los estudiantes? ¿Cómo se puede mejorar la validez de cada pregunta? ¿Hay suficientes datos para decidir cuándo se debe mejorarla y cuándo se debe eliminarla? ¿Se puede aumentar la confiabilidad de la prueba? ¿Cómo variará la confiabilidad del test al disminuir o aumentar su extensión? ¿Cuál fue el índice de facilidad de cada pregunta? ¿Cuál fue el poder discriminativo y la eficiencia de cada opción? ¿Añadido a las respuestas que se dan a estas preguntas, el análisis de la prueba permitiría descubrir, además, otras fallas técnicas específicas.

Este panorama general puede resultar muy amplio a primera vista, sin embargo, el maestro enfrenta siempre el problema de la evaluación, y las alternativas de solución que se le proponen, si se le propone alguna, las considera fuera de su alcance la mayor parte de las veces. En estas condiciones, la tarea del maestro se convierte en una rutina, se pierden muchas horas de trabajo y probablemente también grandes cantidades de materiales de evaluación de buena calidad.

Una prueba preparada por un maestro generalmente se utiliza una sola vez: se aplica, se califica y se desecha.

¿Qué consecuencia tiene esto, además de la pérdida de esfuerzos y recursos ya señalada? La respuesta es clara: el maestro nunca sabrá si su prueba medía lo que él quería medir, habrá asignado calificaciones a sus estudiantes sin poder asegurar qué nivel de conocimientos y habilidades poseían, y habrá desechado preguntas que probablemente podrían haber sido utilizadas nuevamente con un mínimo de trabajo de revisión y reestructuración. Todos estos factores aunados, tienen una repercusión directa sobre la calidad del proceso enseñanza-aprendizaje.

Ante este panorama, y concientes de que existen muchos trabajos de alta calidad en el campo del análisis de ítemes, hemos elaborado el presente manual con el objeto de dar al maestro, en los términos más sencillos posibles, la información y la técnica necesaria para que analice él mismo sus pruebas de aula.

En ningún momento pensamos hacer un aporte de alto nivel en el campo de la evaluación educativa; por el contrario, es nuestro único deseo participar por medio de un instrumento sencillo, en la divulgación de conocimientos técnicos y de procedimientos estadísticos que "hace mucho tiempo" son la herramienta de trabajo de los especialistas, pero cuya repercusión en el hecho educativo, apenas se comienza a sentir, o

prácticamente no se ha sentido, principalmente en los países latinoamericanos.

C. Objetivos

El objetivo que el presente trabajo pretende alcanzar es ofrecer un instrumento que

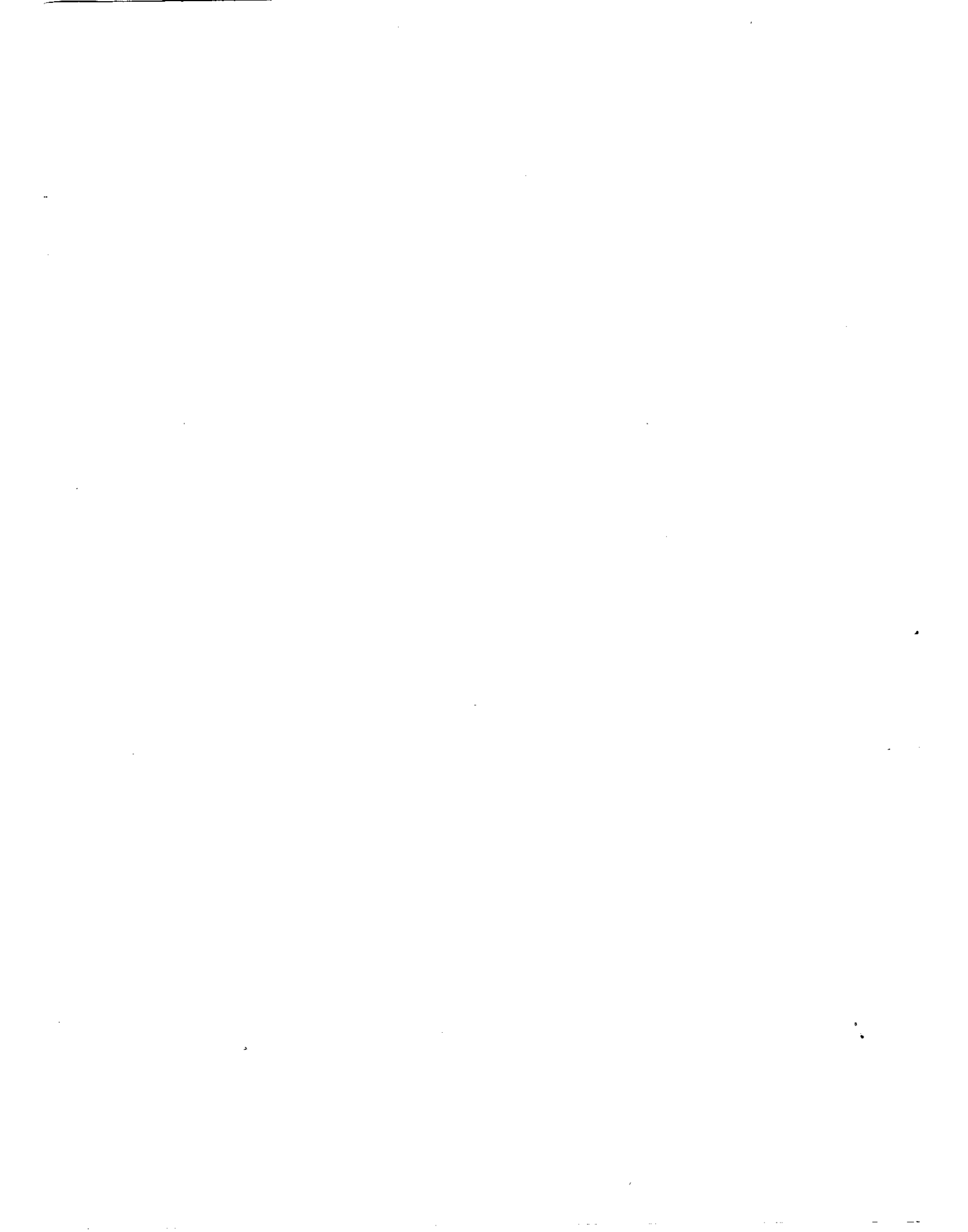
1. Facilite el análisis de pruebas objetivas a nivel de aula.
2. Facilite la evaluación de los resultados del aprendizaje.
3. Estimule al maestro en la construcción de pruebas objetivas confiables, con ítemes que han sido probados previamente y cuyas características principales han sido estadísticamente establecidas y lo instruyen en la forma de hacerlo.
4. Promueva la integración de bancos de ítemes a nivel institucional que faciliten la confección futura de instrumentos de evaluación de buena calidad.
5. Divulgue los conocimientos respecto al análisis de pruebas y que inicie a los maestros en un estudio profundo de estos temas.

D. Definición operacional de términos

A continuación se definen, de una manera sencilla, los términos más usuales en el análisis de ítemes.

1. Análisis de ítemes. Es un procedimiento estadístico que consiste en determinar los índices de facilidad, discriminación y validez de los ítemes que constituyen un test.
2. Confiabilidad. Es la consistencia y constancia de las puntuaciones obtenidas en dos o más mediciones sucesivas, hechas con el mismo instrumento.
3. Correlación. Es la tendencia de una puntuación a variar concomitantemente con otra.
4. Desviación estándar. Medida de la variabilidad o dispersión de un conjunto de puntuaciones, con respecto a la media de esas puntuaciones.
5. Discriminación. Es la capacidad que tiene un ítem de establecer diferencias entre niveles de habilidad.
6. Distractor. Cualquiera de las alternativas que constituyen respuestas incorrectas de un ítem de elección múltiple o de asociación.
7. Facilidad del ítem. Porcentaje de muestras correctas de un ítem.

8. Grupo criterio. Es el conjunto de individuos cuyas respuestas se utilizan para hacer el análisis de ítemes.
9. Ítem. Cada una de las cuestiones, preguntas o problemas que se plantean en un test.
10. Media. Es el resultado de dividir la suma de una serie de datos, entre el número de ellos.
11. Ítem objetivo. Es aquel en cuya respuesta no es posible la divergencia de opiniones.
12. Test objetivo. El que está compuesto de ítemes objetivos.
13. Validez. Es el grado de precisión con que un instrumento mide lo que pretende medir.
14. Varianza. Es el cuadrado de la desviación estándar.



III. REVISION DE LA LITERATURA

Cuando se revisa la literatura relativa al análisis de ítemes, se nota que son muchas las obras que tratan sobre este tema, con enfoques diferentes y con niveles muy diversos. Sin embargo, existen obras representativas a las cuales la mayoría de los autores se refieren en sus trabajos y que constituyen la base sobre la cual se establecen nuevos procedimientos. Una de tales obras es la de Guilford. Este autor, además de sus contribuciones personales al campo de la medición psicológica y educativa, ha hecho la recopilación de procedimientos propuestos por otros autores; por esta razón, consideramos apropiado incluir una pequeña parte de su obra en esta revisión.

Otro autor que, además de los tratamientos tradicionales ofrece en su obra algunas contribuciones novedosas que tienden a enriquecer la información obtenida en un análisis de ítemes es Robert Ebel, cuyo trabajo también comentamos aquí.

Desde luego, no se pretende que en estos breves resúmenes aparezca una revisión exhaustiva de la literatura existente, sino que solamente se desea proporcionar al usuario de este manual algunos antecedentes sobre el análisis de pruebas y facilitar a la persona interesada el estudio profundo de algún tema particular.

Existen muchos otros autores como Magnuson, Garret, Gronlund, Adkins, Thorndike, etc., cuyos trabajos resultarán de gran interés a quienes deseen estudiar a fondo varios procedimientos alternativos para el análisis de ítemes. Para este efecto, consúltese la bibliografía incluida al final de este manual.

Comentario de las obras relevantes

En este punto, se resumen diecinueve métodos para la determinación de la validez de ítemes, tomados de una recopilación hecha por Guilford, 1936.

Es importante hacer la diferencia entre lo que significa la dificultad de un ítem de prueba y su valor como elemento del diagnóstico. Aunque no son independientes estos dos factores, sí son lógicamente diferentes. Consideramos que un ítem tiene un alto poder discriminativo para el diagnóstico si pueden acertar con la respuesta casi todos los estudiantes que tienen habilidad superior a la requerida para responder correctamente el ítem y si además no acierta la mayoría de los que tienen habilidad inferior a esta condición. Por otro lado, si a un ítem contestan en igual forma los individuos de cualquier nivel, se dice que el ítem no tiene poder discriminativo. Se concluye entonces que los ítemes que resulten demasiado difíciles o fáciles no diagnostican diferencias de habilidad. No

son buenos tampoco los ítemes de dificultad media a los cuales responden acertadamente muchos estudiantes de habilidad baja y fallan los de habilidad superior. Teóricamente el nivel de habilidad está en razón directa de la cantidad de aciertos.

La problemática del análisis matemático de ítemes, ha sido obviada por los evaluadores, y una de las razones más poderosas es el desconocimiento o la poca familiaridad con las técnicas psicofísicas necesarias. Binet y sus seguidores adoptaron la edad mental como patrón de habilidad, usaron la proporción de individuos que acertaron en un ítem como indicador de nivel, pero la edad mental nunca ha sido seriamente considerada como una escala racional de habilidad (Guilford; 1936). Otros sustitutos de la edad mental fueron puntajes brutos, clasificación de deciles, centiles u otros. Lógicamente esto correspondería al desconocimiento de algunos principios y técnicas fundamentales para determinar el verdadero valor de diagnóstico del ítem.

Con el objeto de descubrir racionalmente la validez de los ítemes, es necesario seguir criterios si se quiere arbitrarios. En la clasificación que sigue, se han propuesto seis clases de métodos. Esto se basa en el hecho de que los puntajes criterio pueden ser tratados en tres formas

diferentes: los sujetos pueden ser divididos en dos categorías extremas; pueden ser separados en un número de clases superior a dos; o pueden ser tratados como una distribución continua (Guilford; 1936). El ítem, además de ser clasificado de esta manera, puede agruparse en categorías dicótomas o en un número mayor de clases. Si éste es el caso, el puntaje criterio puede ser asimismo dicótomo, en grupos o continuo (Guilford; 1936). Si se trata de una prueba objetiva de selección múltiple, el puntaje es siempre universalmente mantenido en términos de una variable continua.

B. Puntaje dicótomo y grupos de criterio múltiples

1. Método de precisión. Aquí, los sujetos son clasificados en tercios, cuantiles, o cualquier otra agrupación que contenga números iguales de individuos; con el fin de dar una línea base de unidades que correspondan a medidas percentiles en lugar de puntajes brutos. Este método está compendiado en los trabajos de Paterson, Uhrbrock y Richardson. En este método, el valor de diagnóstico del ítem está indicado por la declinación de una curva. La solución gráfica permite determinar el valor discriminativo de cada uno y decidir cuáles son los más discriminativos.

2. Puntaje múltiple y criterio continuo (relación de correlación eta). Con este método, los sujetos que dan

una respuesta a todos los ítemes, o reciben el mismo puntaje ponderado, son separados (Guilford; 1936). La solución que se presenta está determinada por la fórmula:

$$\eta = \frac{\sigma_{y'}}{\sigma_y} = \frac{\text{Valores predichos}}{\text{Valores observados}}$$

para relaciones lineales.

3. Método de McCall. El método utiliza la siguiente fórmula:

$$ADY'' = \frac{\sum [n(Y'' - MY)]}{N}$$

Y'' = Media del puntaje criterio de un grupo

MY = Media del puntaje criterio del grupo total

n = Número de individuos de un grupo particular de respuesta

N = Número total de individuos

Esta fórmula, en el fondo, es una especie de correlación.

4. Método de Long. La fórmula para este método es dada por Barthelmess:

$$L = \frac{(m_1 - m_2)f_1f_2 + (m_1 - m_3)f_1f_3 + \dots + (m_{n-1} - m_n)f_{n-1}f_n}{N\sigma_y}$$

en donde:

m = Media de los puntajes criterio de individuos que dan determinada respuesta

f = Número de individuos que dan esa respuesta

σ_y = Desviación típica del grupo criterio

N = Número total de individuos

5. Método de traslape de Vincent. El método consiste en dividir el grupo criterio en dos partes, los que aciertan y los que fallan. El porcentaje del grupo que falló, que excede la media del grupo que acertó, se toma como medida de validez. Este método (a diferencia de otros), cuando produce un coeficiente alto (Coeficiente de Traslape), indica baja validez y viceversa. Matemáticamente podemos expresar esta afirmación como:

$$VO = pf > M_{dnp}$$

para:

VO = Coeficiente de traslape de Vincent

pf = Proporción del grupo que falló

M_{dnp} = Media del grupo que acertó

Este método también se le acredita a Thorndike.

6. Método de Vincent modificado. Barthelmess perfeccionó el método de traslape de Vincent e incluyó como una innovación a aquellos que acertaron en el ítem y quedaron por debajo de la media del grupo que falló. El coeficiente de traslape será la media de dos proporciones expresada como:

$$BO = \frac{P_{f > M_{dnp}} + P_{p < M_{dnf}}}{2}$$

7. Método de traslape de Long. Long propone un método de traslape que en su opinión es superior al de Vincent y lo expresa matemáticamente como:

$$LO = 1 - \frac{2 \cdot \Sigma (\text{"aciertos" por debajo de Errores})}{p \cdot f}$$

en donde:

LO = Coeficiente de traslape de Long

p = Número total de aciertos

f = Número total de errores

8. Método de traslape ponderado de Long. Long también construyó un instrumento donde incluía el factor de dificultad como indicador del poder de diagnóstico. El poder de discriminación estará determinado por la dificultad relativa para el grupo al que se le aplicó.

La fórmula se reduce a:

$$LWO = PF - 2 \frac{\sum (\text{"aciertos" bajo "errores"})}{(n^2/2)}$$

donde n es el número de ítemes.

9. Método r biserial. Consiste en dividir los puntajes de la prueba en dos categorías en donde las medidas de criterio pueden ser continuas con miras a calcular la conocida r biserial. Esta correlación ha servido como patrón de comparación de los resultados de otros análisis.

10. Método de suma de ajustes de Lentz. Lentz y sus colaboradores proponen un método que es aplicable especialmente a los tests de personalidad, en donde el sujeto responde si o no. Cada ítem es ponderado de acuerdo con la frecuencia con que los otros ítemes son contestados de la misma manera (Guilford; 1936),

El coeficiente de ajuste para el ítem es:

$$CAg = \frac{a_1 + a_2 + a_3 + \dots + a_n}{N}$$

donde:

CAg = Coeficiente de ajuste

a_1 = Número total de respuestas dadas por el sujeto, que concuerdan con la respuesta de ese sujeto al ítem que se está considerando

a_2 = Los mismo para el segundo individuo, y así sucesivamente hasta a_n .

N = Número total de individuos

11. Método de correlaciones sucesivas. Este método ha sido atribuido a Toops y tiene una buena calidad técnica. El procedimiento es el siguiente: primero se separan los ítemes con la validez más alta. En segundo lugar, se para cada uno de estos ítemes con cualesquiera otros. Se correlacionan los puntajes de esta combinación y ésta con la correlación del criterio hasta encontrar el par más válido. Como tercer paso, se establece el tercio, cuarteto o quinteto más válido en orden, constituyendo así la combinación más válida de ítemes.

12. Método de los residuos sucesivos. Este método fue concebido inicialmente por Horst. El indica que la selección de ítemes para el test no es simplemente el problema de seleccionar los que poseen alta validez. Si se trata de confeccionar una batería de tests, la correlación de los tests individuales debe ser baja, si se desea obtener beneficio de la combinación. La técnica entonces, debe

ser la de una especie de correlación múltiple. Horst propone el siguiente procedimiento abreviado: primero, se selecciona el ítem que tenga la correlación más alta con el criterio, se calcula la parte del criterio original que no predice el mejor ítem del tests. Este constituye el primer residuo que pasa a ser el nuevo criterio. Segundo, se correlacionan cada uno de los ítems restantes con el residuo calculado anteriormente. Para calcular cuál predice mejor en forma individual, se establece el residuo no predicho por los ítems no correlacionados y se vuelve a repetir el procedimiento. El trabajo es tedioso y puede continuarse hasta que la mayor parte del criterio original ha sido considerada. El proceso completo puede repetirse dando una serie de ítems que predigan mejor una gran parte del criterio. Podemos concluir que el procedimiento es complejo y requiere fórmulas muy largas para los cálculos básicos.

13. Puntajes dicótomos y grupos criterio dicotomizados.

Cuando se trabaja con ítems con respuestas acertadas o desacertadas, o cualesquiera otras dos categorías y, además, si el grupo criterio se divide en dos (altos y bajos), el indicador más seguro y simple se obtiene calculando el número de aciertos en el grupo alto y bajo. Cada grupo puede dividirse en tercios y cuartos más altos y más bajos de toda la población examinada. El indicador más válido

es la diferencia entre la proporción del grupo alto que acierta el ítem y la proporción del grupo bajo. Este método es una muy buena aproximación a otros que requieren mucho más tiempo.

14. Método de relación crítica. El método consiste en buscar los dos grupos extremos y establecer las proporciones de aciertos en cada uno de ellos con los respectivos errores estándar y errores probables. A continuación se calculan las diferencias entre las proporciones y la desviación típica (sigma) o el error probable (K) de la diferencia, por medio de fórmulas especiales. El valor de la relación que se toma como referencia, es un tanto arbitrario y se considera como significativos aquéllos que oscilen entre $d/\sigma d = 2$ y $d/Kd = 3$. Dado que los valores de σp o Kp dependen de N, el nivel requerido de significación se puede obtener fácilmente si se analiza un gran número de casos. Si existe una N pequeño, los mismos ítemes pueden no requerir el mismo nivel de significación. Este método difiere del de traslape puesto que aquel es independiente del N utilizado.

15. Índice de validez de Clark. Clark propone la fórmula:

$$I.V. = \frac{P - D}{I - D}$$

donde:

D = Proporción de individuos que fallaron el ítem

P = Proporción del grupo bajo que falló el ítem

I.V. = Índice de validez

16. Índice de discriminación de Cook (A). La labor de Cook en este aspecto, fue la de encontrar índices de validez que tuvieran como base la validación de ítems de tests de deletreo. Su método consistió en calcular el porcentaje de estudiantes que deletrean mal una palabra dentro de un grupo de alumnos formado por todos aquéllos que deletrearían mal si la palabra fuera perfecta en su poder de discriminación.

Deduca la fórmula:

$$IDa = \frac{F}{E'}$$

donde:

E' = Parte más baja del grupo criterio (igual al número a E), cuando E es el número total de alumnos que falló en el ítem.

F = Número de alumnos del grupo E que realmente falló en el ítem.

El índice A es, además, el porcentaje de los estudiantes que fallaron sobre los que se esperaba que fallaran suponiendo una discriminación perfecta para el ítem.

17. Índice de discriminación de Cook. (B). Este índice se define como el porcentaje de alumnos que delectran mal una palabra en un grupo formado por todos los alumnos que lo hacen, si la palabra fuera perfecta en su poder de discriminación, después de que aquellos alumnos que fallaron en la palabra en un grupo normal del mismo tamaño, han sido restados.

El índice está dado por:

$$IDb = \frac{FN - E^2}{EN - E^2}$$

Donde:

E = Número de alumnos del grupo total que falló en el ítem

F = Número de miembros del grupo "E" (grupo que se esperaría que fallara si el ítem fuera perfecto) que falló en el ítem

N = Número de individuos.

18. Índice de discriminación de Cook (C).

$$IDc = \frac{P_1 - P_4}{75}$$

Donde:

P_1 = Proporción de aciertos en un ítem entre los del cuartil más bajo

P_4 = Proporción que acierta en el cuartil más bajo
 75 = Número de percentiles de habilidad entre los dos
 cuartiles extremos

En resumen, el índice C no es otra cosa que la pendiente de una línea que une dos porcentajes de los cuartiles más altos y más bajos de habilidad cuando las unidades de precisión y habilidad se consideran iguales.

19. Índice de discriminación de Cook (D). Es una proporción entre el cuarto superior de habilidad y el cuarto más bajo:

$$IDd = \frac{P_1}{P_4}$$

Donde:

P_1 = Proporción que acierta en el ítem (cuartil superior)

P_4 = Proporción que acierta en el ítem (cuartil inferior)

VI. MANUAL DE TRABAJO

Para el análisis de ítemes se confeccionaron dos formatos de trabajo con indicaciones precisas a fin de que el educador pueda fácilmente realizar la tarea.

Para resolver algunas inquietudes particulares se presenta la fundamentación teórica de algunos conceptos, fórmulas y decisiones.

Aún cuando esta información no es un requisito indispensable para poder desarrollar el análisis propuesto, su lectura aclarará algunos conceptos importantes.

A. Fundamentos teóricos

1. Valor del análisis de ítemes. No queda duda que la mejor manera de hacer el análisis de un test, a fin de mejorar su calidad, es basarse en las respuestas que da cada uno de los examinados.

Dicho análisis puede indicarnos cuáles ítemes fueron fáciles, cuales difíciles, qué razones pudieron influir en la no selección de una respuesta correcta, porqué no funcionó como se esperaba y cómo puede ser mejorado. Hay mayor certeza de que un test arroje mejores resultados si sus ítemes han sido previamente analizados que cuando se aplican por vez primera. En la medida que un maestro pruebe sus ítemes,

adquirirá mayor destreza en la confección de buenas pruebas y colateralmente aumenta la confiabilidad de sus exámenes.

2. Proceso general del análisis de ítems. El proceso comienza después que las pruebas objetivas hayan sido debidamente calificadas y asignadas sus respectivas puntuaciones. Son muchos los procedimientos que han usado los maestros a través del tiempo y que pueden resumirse, en términos generales, en seis pasos:

- a. Ordenamiento de los puntajes de mayor a menor.
- b. Establecer el grupo criterio, o sea, separar el 27 por ciento de los exámenes con los puntajes más altos e igual número del grupo más bajo.^{/1/}
- c. Contar el número de veces que cada alternativa fue seleccionada en el grupo alto y en el grupo bajo.
- d. Registrar esta cuenta en una hoja especial.
- e. Adicionar los resultados de la respuesta correcta del grupo superior y del inferior. Dividir esta suma entre la suma máxima posible (suma total del grupo

^{/1/} Véase el inciso c) de este capítulo para una explicación completa.

superior más el inferior). Multiplicar la fracción decimal por 100. Este resultado será el índice de facilidad del ítem.

f. Restar la cantidad de respuestas correctas del grupo bajo de la cantidad de respuestas correctas del grupo alto. Dividir esta diferencia entre la diferencia más grande posible, esto es, el número de personas del grupo superior o del inferior. Este cociente, expresado como una fracción decimal, es el índice de discriminación.

3. Selección de los grupos criterio. El seleccionar el 27 por ciento alto y bajo para un análisis de ítemes como grupo criterio, no corresponde a una concepción arbitraria. T. Kelley demostró que esta proporción (27 por ciento) provee la mejor relación entre los siguientes objetivos deseables: hacer los grupos extremos tan grandes como sea posible, pero hacerlos a la vez tan diferentes como sea posible. Lo que Kelley demostró fue que tomando grupos superiores o inferiores del 27 por ciento del grupo total, se puede decir con gran confianza que los del grupo alto son superiores en la habilidad medida por el test, a los del grupo inferior (R. Ebel; 1954). Si por ejemplo, tomáramos como grupo criterio un 10 por ciento del total, la diferencia en niveles promedio de habilidad podría ser más grande por lo reducido

del grupo, que en uno de 27 por ciento; pero por otro lado podemos estar menos seguros de que sus niveles promedio de habilidad en los dos grupos sean menos confiables.

Si tomásemos el tamaño máximo para el análisis (50 por ciento) como grupo criterio, no tendríamos posibilidad de tener una zona de discriminación, lo que daría como resultado la poca certeza de que efectivamente todos los estudiantes del grupo bajo pertenezcan al 50 por ciento inferior y no exista traslape en las dos categorías.

La conclusión a la cual arribamos es que con el 27 por ciento como grupo criterio se conjugan tamaño adecuado y diferencias óptimas en niveles de habilidad.

4. Grupos criterio especiales. Mientras más grande sea el número de individuos que forman el 27 por ciento, tanto más confiables serán los resultados del análisis. Cuando el análisis se hace a nivel de programas generales de evaluación, el 27 por ciento viene a representar varios cientos de casos. Si se da la situación de que el número de casos a analizar excede a las posibilidades físicas de análisis, es conveniente entonces (en grupos muy numerosos) escoger un porcentaje menor. Ante esta alternativa es recomendable escoger al azar un número más pequeño para cumplir con los preceptos iniciales: mientras más grande sea la

diferencia entre los grupos, el nivel promedio de habilidad es mejor y se obtiene además, un grupo de razonable tamaño. Un estudio de Asche demostró que la confiabilidad de los índices de discriminación obtenidos para los grupos extremos superiores e inferiores al 10 por ciento, tendían a ser más grandes que el obtenido para grupos extremos superiores e inferiores de 27 por ciento, cuando los grupos extremos incluían un ciento de estudiantes cada uno (Robert Ebel; 1954). Cabe anotar que los resultados finales para cada ítem no son totalmente comparables cuando se toman muestras de 27 por ciento y 10 por ciento. Los índices para el grupo criterio del 10 por ciento tienden a ser sensiblemente más grandes que los calculados con un 27 por ciento. Se recomienda entonces verificar o normalizar el por ciento escogido para el grupo criterio.

Flanagan ha sugerido que el cómputo de las respuestas a los ítemes y los índices de discriminación obtenidos pueden hacerse más confiables si se aumenta el peso de las respuestas dadas para estudiantes cuyos puntajes criterios fueron extremadamente altos o bajos. Una de las sugerencias que él ofrece cuando los grupos son muy pequeños, es que las respuestas del 9 por ciento inferior y superior (extremas) sean contadas dos veces con las respuestas del 20 por ciento siguiente, pero contadas éstas últimas sólo una vez. Esto tiene las siguientes consecuencias deseables:

- a. Incrementar la diferencia en los niveles promedios de habilidad.
- b. Incrementar el tamaño de los grupos criterio superior e inferior.

Construyó Flanagan tablas que ayudan a calcular estos índices de correlación por el método ponderado.

5. Cuenta de las respuestas. Contar las respuestas a cada ítem es el trabajo más tedioso y dilatado del análisis de ítemes; sin embargo, con un poco de ingenio y organización del trabajo se puede salvar esta situación.

6. Índice de discriminación. Generalmente se representa simbólicamente por la letra D y es de fácil interpretación y cálculo. El valor esperado para D para ítemes de 50 por ciento de facilidad son mayores que para ítemes de un 40 por ciento o 60 por ciento de facilidad; pero estos a su vez son más grandes que para ítemes con cero o 100 por ciento de facilidad. La facilidad óptima para cada ítem depende de otras características colaterales del ítem y del test.

Un criterio diferencial que proporciona alguna información, es el examinar si un sujeto acierta en un ítem mientras que otro falla; cosa que no sucedería cuando se da el

caso de que ambos acierten o ambos fallen en él. Por ejemplo, si los alumnos A, B, C y D contestan correctamente a un ítem, pero los alumnos E y F fallan, nos proporcionan ocho elementos de información diferencial. Tan pronto como son comparados, el alumno A muestra más habilidad que el E y que el F, y así sucesivamente. En resumen, el número de elementos de información diferencial que un ítem puede brindar es el producto del total de los que acierten por el número de los que fallan en él.

Si el objetivo primordial del análisis de ítemes es aumentar la confiabilidad del test, se seleccionarán entonces los ítemes que tengan la más alta discriminación. Debe entenderse entonces que la facilidad necesaria del ítem no es directamente proporcional al valor del ítem dado el caso de que ningún ítem demasiado fácil o difícil puede arrojar buena discriminación si tomamos como criterio la diferencia entre grupo superior e inferior.

No todas las discriminaciones que un ítem proporciona son totalmente confiables. Si un ítem es contestado por un 50 por ciento de los buenos estudiantes y por un 50 por ciento de estudiantes deficientes, proveen exactamente las mismas discriminaciones correctas e incorrectas. Será diferente la discriminación cuando se trate de estudiantes que

contestan correctamente un ítem en contraste con los estudiantes deficientes que fallan en él. Se consideran estas discriminaciones correctas. Por el contrario, será una discriminación incorrecta cuando los estudiantes buenos fallan en un ítem mientras que los estudiantes deficientes aciertan.

Los índices de discriminación de los ítemes de todo tipo, están sujetos a un considerable error de muestreo (Ebel; 1954). Mientras menor sea el número de hojas de respuesta considerado en el análisis, más grande es el error de muestreo. Un ítem que representa una alta discriminación en muestras pequeñas puede resultar con baja o negativa discriminación en muestras grandes. Lo anterior no significa que prescindamos de hacer un análisis de ítemes cuando las muestras son pequeñas; sólo, que debemos tomar con algunas reservas sus resultados.

En resumen, el grado de calidad de un test depende de qué tan grandes son las muestras analizadas y qué tan pequeños los errores de muestreo.

7. Criterio para el análisis de ítemes. Para el análisis de ítemes se hace uso de un criterio interno para la selección de grupos de alto y bajo rendimiento. El puntaje total del test cuyos ítemes van a ser analizados se

usa como criterio en lugar de alguna otra medida independiente (externa) del rendimiento.

Para poder afirmar que un ítem muestra una alta discriminación y es bueno, cabe suponer que el test, del cual el ítem forma parte, es en general un buen test. Se debe admitir además que el análisis de ítems que usa un criterio interno, puede mejorar a un test solamente en aquello que pretenda medir. Se podía pensar que un criterio externo puede ser mejor, sin embargo, Ryans estableció que los criterios externos disponibles son de valor limitado en la validación de ítems para pruebas de información profesional (Ebel; 1954).

El uso del puntaje total del test como criterio para la selección de grupos alto y bajo tiene dos ventajas indiscutibles: los límites de habilidad que se establecen, responden a la experiencia del constructor del test, y la medida referencial (puntaje total del test) siempre está disponible.

B. Descripción de formatos de trabajo e instrucciones para su manejo

Las descripciones e instrucciones que se dan a continuación, se refieren a las características de los formatos de trabajo que se han diseñado para realizar el análisis de

ítemes y pruebas y a la forma de utilizarlos. Se recomienda al lector, referirse al apéndice para lograr una clara comprensión de estos formatos.

Descripción del formato de trabajo No. 1

Este formato ha sido diseñado para obtener los índices de facilidad y discriminación de cada uno de los ítemes que constituyen el test; asimismo un análisis de las alternativas y un cuadro de decisiones que permitirá determinar si se va a conservar o a rechazar el ítem.

Las filas contienen los datos relativos a los ítemes y las columnas numeradas (1, 2, 3, ...) contienen los datos relativos a estudiantes.

Cada casilla ha sido dividida en cuatro secciones. El diagrama No. 1 ejemplifica la colocación de los resultados del alumno uno del grupo superior, el alumno uno del grupo inferior respecto al ítem número uno

Alumno	Item		
	1	2	3
1	S	W	X
	I	Y	Z

Alumno	Item		
	1	2	3
1	S	1	
	I	O	B

Diagrama No. 1

Las cuatro secciones se llenarán de la siguiente manera:

Sección W. Se llenará con uno (1) si la respuesta del estudiante superior es correcta, con cero (0) si es incorrecta. En caso de no existir respuesta (omisión) se colocará un asterisco (*).

Sección X. Si la sección W es uno (1), esta sección X no se usa, dado que el estudiante ha acertado en el ítem. Si la sección W es cero (0) debe colocarse aquí la letra seleccionada por el estudiante en la hoja de respuestas.

Sección Y. Se sigue el mismo procedimiento que para la sección W, teniendo el cuidado de tomar como estudiante uno, al que obtuvo el mayor puntaje en el grupo inferior.

Sección Z. Sígase el mismo procedimiento que en la sección X, sólo que con el grupo inferior.

Este procedimiento se repite hasta haber considerado todos los ítems y todos los estudiantes del grupo criterio.

En las columnas restantes se realizan los cálculos de los niveles de facilidad y discriminación, y para la cuenta del número de alternativas incorrectas seleccionadas para cada ítem. La sección final del formato es un cuadro de decisiones en donde se resumen los resultados obtenidos.

1. Instrucciones para el manejo del formato de trabajo
No. 1.

a. De acuerdo con el número total de estudiantes examinados (N) y utilizando la Tabla No. 1, ubique en ella el valor correspondiente al tamaño de los grupos superior e inferior para $n = 27$ por ciento de N .

b. Ordene todas las hojas de respuesta en orden decreciente de puntajes.

c. Sin alterar el orden de las hojas de respuesta, separe tantas hojas cuantas indique el cálculo del 27 por ciento (grupo superior).

d. A partir de la última hoja y sin alterar el orden de ellas, separe la misma cantidad que para el grupo superior. Estas constituirán el grupo inferior ($n = 27$ por ciento de N).

e. Llene con unos y ceros la primera parte del Formato de Trabajo No. 1, de acuerdo con las instrucciones que se dieron en la descripción de éste.

f. Obtenga la suma horizontal de los valores contenidos en la sección X de cada casilla (grupo superior). Coloque el resultado en la columna indicada con la palabra suma.

g. Proceda de igual forma para el grupo inferior (Sección Y).

h. Obtenga $S - I$ restando el valor de la parte inferior de la casilla de la columna SUMA, al valor de la parte superior de esta casilla. Coloque el resultado en la columna $S - I$ correspondiente.

i. Divida cada uno de los valores $S - I$ entre el tamaño del grupo superior (n) y coloque el resultado en la parte superior de la casilla correspondiente de la columna $\frac{S - I}{n} = D$. Estos son los índices de discriminación para cada ítem.

j. Calcule las $S + I$ adicionando los valores correspondientes a cada casilla de la columna suma. Escriba los resultados en las casillas correspondientes.

k. Divida cada $S + I$ entre el tamaño del grupo criterio ($2n$) y escriba los resultados en la parte superior de la casilla correspondiente. Estos valores serán los niveles de facilidad de cada ítem.

l. Con el valor D obtenido y consultando la tabla de decisiones No. 6, verifique por comparación si cada ítem discrimina o no y escriba la clave correspondiente, según el caso, en la parte inferior de la casilla de la columna $D = \frac{S - I}{n}$.

m. Compare los valores F obtenidos para cada ítem, con los que proporciona la Tabla No. 6 y verifique su nivel de facilidad. Escriba en la parte inferior de la casilla correspondiente de la columna F, de acuerdo con la clave encontrada en la tabla.

2. Suma de las respuestas erróneas seleccionadas

a. Cuente horizontalmente el número de letras (A, B, C, D, E) que aparezcan para cada ítem tanto en el grupo superior como en el inferior y coloque los resultados numéricos en la parte de la casilla correspondiente. Repita el procedimiento hasta haber agotado todos los ítems.

b. Consulte la tabla de decisiones y revise los resultados. Llene la columna de decisiones en el Formato de Trabajo.

3. Instrucciones para el manejo del formato de Trabajo

No. 2.

a. Tomando como base las hojas de respuestas, llene las columnas 1, 2, 3, etc. con uno (1) si el sujeto considerado acertó en la respuesta y con cero (0) si falló. Repita el procedimiento para cada uno de los estudiantes del grupo criterio (2n).

- b. Calcule las sumas horizontales para cada uno de los estudiantes, y coloque el resultado en la casilla PB (puntaje bruto).
- c. El valor de las columnas 1PB, 2PB, 3PB...etc. se encuentra multiplicando cada uno de los valores de las columnas 1, 2, 3,... etc. por el valor PB obtenido para cada fila.
- d. Obtenga las sumas verticales para cada ítem (columnas 1, 2, 3...) y escriba los resultados en la fila S (suma).
- e. Obtenga las sumas de las columnas 1PB, 2PB, 3PB... y divida cada resultado entre el valor de S (sumas) correspondiente. Escriba los resultados en la fila MI, en la celda correspondiente.
- f. Calcule el valor PBT sumando los valores de la columna PB. Verifique este valor obteniendo la suma horizontal de la fila S.
- g. Calcule los valores PI, dividiendo cada uno de los valores de la fila S, entre el $2n$ calculado, o bien, transcriba los índices de facilidad (F) calculados anteriormente.
- h. Obtenga MT sumando todos los valores de la fila PI. Este valor es la media total de la distribución.

- i. Calcule los valores $(PB - MT)$, restando cada valor de la columna PB el valor MT (media total).
- j. Eleve al cuadrado cada valor $(PB - MT)$, utilizando la Tabla No. 2 y escríbalo en la casilla correspondiente.
- k. Calcule la suma de la columna $(PB - MT)^2$ y escríbalo en el lugar correspondiente en la fila S.
- l. Obtenga ST^2 (varianza total), dividiendo la suma obtenida en el paso anterior entre el número de sujetos del grupo criterio $(2n)$.
- m. Usando la Tabla No. 2 extraiga la raíz cuadrada de ST^2 para obtener la ST (desviación estándar).
- n. Calcule los valores de la fila DM (diferencia de medias), restando MT a cada valor de MI y escribiendo el resultado en la celda correspondiente.
- ñ. Calcule los valores de la fila $\frac{DM}{ST}$ dividiendo cada valor de DM entre ST y escriba el resultado en la casilla correspondiente.
- o. Obtenga el valor de $\sqrt{\frac{PI}{QT}}$ en la Tabla No. 3, tomando como valor de entrada el PI calculado para cada ítem y multiplíquelo por el valor de $\frac{DM}{ST}$. El resultado

escribalo en la casilla correspondiente de la fila r_{pbis} . Los valores de la fila r_{pbis} , son las correlaciones punto biserial de cada uno de los ítemes con respecto al test y constituyen una medida de la validez de cada ítem.

C. Presentación de fórmulas estadísticas

A continuación se hace una breve descripción de las fórmulas estadísticas utilizadas en este trabajo.

1. Índice de facilidad

$$F = \frac{S + I}{2n}$$

donde: F = Índice de facilidad de un ítem particular expresado en forma decimal

S = Número de respuestas correctas dadas por el grupo superior al ítem considerado

I = Número de respuestas correctas dadas por el grupo inferior al ítem considerado

2n = Tamaño del grupo criterio

Nota: si se desea expresar F en porcentaje, multiplique por cien.

2. Índice de discriminación

$$D = \frac{S - I}{n}$$

donde: D = Índice de discriminación de un ítem particular
expresado en forma decimal

S e I representan lo mismo que en la fórmula del índice de facilidad

n = Tamaño del grupo superior o inferior, o sea 27 por ciento del grupo total.

3. Media total de la distribución de puntajes

$$MT = \sum PI$$

donde: MT = Media total de la distribución

PI = Suma de las proporciones de estudiantes que contestaron correctamente cada ítem.

4. Desviación estándar de la distribución de puntajes

$$ST = \sqrt{\frac{(MT - PB)^2}{2n}}$$

donde: ST = Desviación estándar de la distribución de puntajes

MT = Media total de la distribución de puntajes

$2n$ = Tamaño del grupo criterio

5. Correlación punto biserial de cada ítem con el test.

(Medida de validez del ítem)

$$r_{pbis} = \frac{MI - MT}{ST} \sqrt{\frac{PI}{QI}}$$

- donde: r_{pbis} = Correlación punto biserial del ítem con respecto al test
- MI = Media particular de cada ítem, o sea, la media de los PB de los estudiantes que contestaron bien al ítem
- MT = Media total de la distribución de puntajes
- ST = Desviación estándar de la distribución de puntajes
- PI = Proporción de estudiantes que resolvieron correctamente el ítem
- QI = Proporción de estudiantes que fallaron en el ítem

6. Confiabilidad

$$r_{KR-20} = \frac{K}{K-1} \frac{ST^2 - \sum PI QI}{ST^2}$$

- donde: r_{KR-20} = Coeficiente de confiabilidad del test calculado con la fórmula 20 de Kuder-Richardson.
- K = Número de ítems del test
- ST^2 = Varianza de la distribución de puntajes (cuadrado de la desviación estándar)
- $\sum PI QI$ = Suma de los productos de las proporciones de alumnos que acertaron y de los que fallaron en el ítem

Nota: esta fórmula la utilizaremos para calcular el índice de confiabilidad del test, una vez terminado el análisis de ítemes en los formatos de trabajo 1 y 2. Utilice para el cómputo de la confiabilidad, las tablas 3 y 4.

Cuando el grupo criterio ($2n$) sea mayor de 20 sujetos, utilice formatos adicionales teniendo cuidado en el manejo de los subtotales.

V. TRABAJO DE CAMPO

A fin de obtener la opinión y evaluación (en cierta forma) del manual de trabajo, por aquellas personas que lo utilizaron en el análisis de un test, se elaboró una encuesta (ver Apéndice) cuyos resultados se tabulan a continuación:

Item	1	Item	2	3
Respuesta		Respuesta		
Muy útil	2	Muchas veces	2	0
Util	8	Pocas veces	3	3
Foco útil	0	Nunca	5	7

Item	4	5	6	7	11	12	13	14
Respuesta								
Sí	10	10	10	10	10	10	10	10
No	0	0	0	0	0	0	0	0

Tabla número	Item	8	9
	Respuesta		
1	Si	10	10
	No	0	0
2	Si	10	10
	No	0	0
3	Si	10	10
	No	0	0
4	Si	10	10
	No	0	0
5	Si	10	10
	No	0	0

Tabla número	Item respuesta	10
1	Mucho	10
	Poco	0
	Nada	0
2	Mucho	10
	Poco	0
	Nada	0
3	Mucho	10
	Poco	0
	Nada	0
4	Mucho	10
	Poco	0
	Nada	0
5	Mucho	10
	Poco	0
	Nada	0

VI. CONCLUSIONES Y RECOMENDACIONES

A. Conclusiones

Dado el corto tiempo con que se contó para la elaboración de este trabajo y la imposibilidad de hacer una experimentación amplia con él, se formulan las siguientes conclusiones, haciendo notar las limitaciones que lo anterior impone:

1. Las respuestas obtenidas indican que sí es posible poner al alcance del maestro, en corto tiempo, toda la información estadística necesaria para hacer un análisis de pruebas objetivas.

2. La mayor parte de los maestros encuestados no realizan análisis de ítemes.

Si consideramos que los maestros encuestados forman parte de un grupo que ha tenido opción a realizar estudios más avanzados que los de la gran mayoría, además de trabajar en el medio urbano, en donde las oportunidades de desarrollo son mayores, se puede afirmar que la situación planteada inicialmente es más aguda en el medio rural y que se generaliza a nivel nacional.

3. Con base en los resultados de la encuesta, podemos concluir que el análisis de pruebas objetivas es posible, aunque el conocimiento estadístico que tenga el maestro sea muy pobre.

B. Recomendaciones

1. Recomendamos se hagan estudios que involucren una evaluación de este manual, de su aplicabilidad, utilidad y confiabilidad de los resultados a fin de poder utilizarlo en forma general.
2. Se recomienda realizar estudios tendientes a desarrollar manuales semejantes a éste, que sirvan para el análisis de otro tipo de pruebas que las aquí consideradas.
3. Si se desea utilizar este manual con profesores o con estudiantes de magisterio cuyos conocimientos sobre evaluación y estadística no sean muy amplios, deberán hacerse las modificaciones pertinentes para que resulte un instrumento adecuado a las características académicas y socioculturales de tales personas.
4. Puede utilizarse también como un auxiliar didáctico en el entrenamiento rápido de personal.

VII. BIBLIOGRAFÍA

- Adams, G. S. Medición y evaluación en educación, psicología y "Guidance". Barcelona (España), Editorial Herder 1970.
- Ebel, R. Measuring educational achievement. Englewood Cliffs, New Jersey (Estados Unidos), Prentice Hall, Inc., 1965.
- Garret, H. C. Estadística en psicología y educación. Buenos Aires (Argentina), Editorial Paidós, 1966.
- Gronlund, N. E. Elaboración de tests de aprovechamiento. México, Editorial Trillas, 1972.
- Guilford, J. P. Psychometric methods. New York (Estados Unidos), McGraw Hill Book Company, 1936.
- Gulliksen, H. Theory of mental tests. New York (Estados Unidos), Wiley, 1950.
- Lindeman, R. H. Tratado de medición educacional. Buenos Aires (Argentina), Editorial Paidós, 1971.
- Macintosh, H. C. and Morrison, R. B. Objective testing. Londres (Inglaterra), Unibooks, University of London Press, Ltd., 1972.
- Magnuson, D. Teoría de los tests. México, Editorial Trillas, 1972.
- Thorndike, R. L. y Hagen, E. Tests y técnicas de medición en psicología y educación. México, Editorial Trillas, 1973.

APÉNDICE A

ENCUESTA DE EVALUACIÓN

INFORMACION GENERAL

1. En su opinión, el análisis de ítemes a nivel de aula, como un medio para mejorar el proceso enseñanza-aprendizaje, es un método:

Muy útil____ Util____ Poco útil____

2. ¿Ha tenido Ud. experiencias previas en el análisis de ítemes objetivos?

Muchas veces____ Pocas veces____ Nunca____

3. Utiliza Ud. los conocimientos estadísticos involucrados en este aspecto de la evaluación?

Frecuentemente____ Pocas veces____ Nunca____

EVALUACION DEL MANUAL

4. ¿Le parecen claras las instrucciones que se dan para el manejo del Formato de Trabajo No. 1?

SI____ NO____

5. ¿Le parece funcional el diseño del Formato de Trabajo No. 1?

SI____ NO____

Anote algunas sugerencias que considere importantes:_____

6. ¿Le parecen claras las instrucciones que se dan para el manejo del Formato de Trabajo No. 2?

SI____ NO____

7. ¿Le parece funcional el diseño del Formato de Trabajo número 2?

SI _____ NO _____

Anote alguna sugerencia que considere importante: _____

8. ¿Son claras las instrucciones para el manejo de la tabla:

No. 1	SI _____	NO _____
No. 2	SI _____	NO _____
No. 3	SI _____	NO _____
No. 4	SI _____	NO _____
No. 5	SI _____	NO _____

9. ¿Le parece práctico el diseño de la tabla:

No. 1	SI _____	NO _____
No. 2	SI _____	NO _____
No. 3	SI _____	NO _____
No. 4	SI _____	NO _____
No. 5	SI _____	NO _____

Sugiera alguna modificación que crea conveniente: _____

10. ¿Simplifica el trabajo el empleo de la tabla:

No. 1	SI _____	NO _____
No. 2	SI _____	NO _____
No. 3	SI _____	NO _____
No. 4	SI _____	NO _____
No. 5	SI _____	NO _____

11. ¿Le parecen claras las instrucciones para el manejo de la tabla de decisiones No. 6?

SI _____ NO _____

12. ¿Considera que la tabla de decisiones No. 6 es práctica?

SI _____ NO _____

Sugiera alguna modificación que considere importante: _____

13. ¿Usaría Ud. en forma regular este manual en el análisis de sus pruebas objetivas?

SI _____ NO _____

14. ¿Recomendaría Ud. a otros maestros la aplicación de este manual?

SI _____ NO _____

Sugiera alguna modificación que Ud. crea necesaria para mejorar este manual de análisis de ítemes objetivos: _____

APÉNDICE B

Tabla No.1

DETERMINACION DE TAMAÑO DE n

N	n	N	n
39 - 42	11	113 - 116	31
43 - 46	12	117 - 120	32
47 - 49	13	121 - 124	33
50 - 53	14	125 - 127	34
54 - 57	15	128 - 131	35
58 - 61	16	132 - 135	36
62 - 64	17	136 - 138	37
65 - 68	18	139 - 143	38
69 - 72	19	144 - 146	39
73 - 75	20	147 - 149	40
76 - 79	21	150 - 153	41
80 - 83	22	154 - 157	42
84 - 87	23	158 - 161	43
88 - 90	24	162 - 164	44
91 - 94	25	165 - 168	45
95 - 98	26	169 - 172	46
99 - 101	27	173 - 175	47
102 - 105	28	176 - 179	48
106 - 109	29	180 - 183	49
110 - 112	30	184 - 185	50

Nota: en esta tabla:

N = número de alumnos que presentan el test.

n = 27 por ciento de N (grupo superior e inferior tomados en consideración para el análisis).

APÉNDICE C

Tabla No.2

CUADRADOS DE NÚMEROS Y RAÍCES DE CERO A CINCUENTA

Número	Cuadrado	Raíz cuadrada	Número	Cuadrado	Raíz
0					
0	0	0.000	25	625	5.000
1	1	1.000	26	676	5.099
2	4	1.414	27	729	5.196
3	9	1.732	28	784	5.292
4	16	2.000	29	841	5.385
5	25	2.236	30	900	5.477
6	36	2.449	31	961	5.568
7	49	2.646	32	1024	5.657
8	64	2.828	33	1089	5.745
9	81	3.000	34	1156	5.837
10	100	3.162	35	1225	5.916
11	121	3.317	36	1296	6.000
12	144	3.464	37	1369	6.083
13	169	3.606	38	1444	6.164
14	196	3.742	39	1521	6.245
15	225	3.873	40	1600	6.325
16	256	4.000	41	1681	6.403
17	289	4.123	42	1764	6.481
18	324	4.243	43	1849	6.557
19	361	4.358	44	1936	6.633
20	400	4.472	45	2025	6.708
21	441	4.583	46	2116	6.782
22	484	4.690	47	2209	6.856
23	529	4.796	48	2304	6.928
24	576	4.899	49	2401	7.000
			50	2500	7.071

APÉNDICE D

TABLA PARA LA DETERMINACION DE LA RELACION DE PROPORCIONES $\sqrt{PI/QI}$
 PARA EL CALCULO DE LAS CORRELACIONES BISERIALES ITEM-TEST

PI	QI	$\sqrt{PI/QI}$	PIQI	PI	QI	$\sqrt{PI/QI}$	PIQI
0.00	1.00	0.0000	0.0000	0.51	0.49	1.0202	0.2499
0.01	0.99	0.1005	0.0099	0.52	0.48	1.0408	0.2496
0.02	0.98	0.1429	0.0196	0.53	0.47	1.0619	0.2491
0.03	0.97	0.1759	0.0291	0.54	0.46	1.0835	0.2484
0.04	0.96	0.2041	0.0384	0.55	0.45	1.1055	0.2475
0.05	0.95	0.2294	0.0475	0.56	0.44	1.1282	0.2464
0.06	0.94	0.2526	0.0564	0.57	0.43	1.1513	0.2451
0.07	0.93	0.2744	0.0615	0.58	0.42	1.1751	0.2436
0.08	0.92	0.2949	0.0736	0.59	0.41	1.1996	0.2419
0.09	0.91	0.3145	0.0819	0.60	0.40	1.2247	0.2400
0.10	0.90	0.3333	0.0900	0.61	0.39	1.2506	0.2379
0.11	0.89	0.3516	0.0979	0.62	0.38	1.2773	0.2356
0.12	0.88	0.3693	0.1056	0.63	0.37	1.3049	0.2331
0.13	0.87	0.3866	0.1131	0.64	0.36	1.3333	0.2304
0.14	0.86	0.4035	0.1204	0.65	0.35	1.3628	0.2275
0.15	0.85	0.4201	0.1275	0.66	0.34	1.3933	0.2244
0.16	0.84	0.4364	0.1344	0.67	0.33	1.4249	0.2211
0.17	0.83	0.4526	0.1411	0.68	0.32	1.4577	0.2176
0.18	0.82	0.4685	0.1476	0.69	0.31	1.4919	0.2139
0.19	0.81	0.4843	0.1539	0.70	0.30	1.5275	0.2100
0.20	0.80	0.5000	0.1600	0.71	0.29	1.5647	0.2059
0.21	0.79	0.5156	0.1659	0.72	0.28	1.6036	0.2016
0.22	0.78	0.5311	0.1716	0.73	0.27	1.6443	0.1971
0.23	0.77	0.5465	0.1771	0.74	0.26	1.6871	0.1924
0.24	0.76	0.5620	0.1824	0.75	0.25	1.7321	0.1875
0.25	0.75	0.5774	0.1875	0.76	0.24	1.7795	0.1824
0.26	0.74	0.5927	0.1924	0.77	0.23	1.8297	0.1771
0.27	0.73	0.6082	0.1971	0.78	0.22	1.8829	0.1716
0.28	0.72	0.6236	0.2016	0.79	0.21	1.9396	0.1659
0.29	0.71	0.6391	0.2059	0.80	0.20	2.0000	0.1600
0.30	0.70	0.6547	0.2100	0.81	0.19	2.0647	0.1539
0.31	0.69	0.6703	0.2139	0.82	0.18	2.1344	0.1476
0.32	0.68	0.6860	0.2176	0.83	0.17	2.2096	0.1411
0.33	0.67	0.7018	0.2211	0.84	0.16	2.2913	0.1344
0.34	0.66	0.7177	0.2244	0.85	0.15	2.3805	0.1275
0.35	0.65	0.7338	0.2275	0.86	0.14	2.4785	0.1204
0.36	0.64	0.7500	0.2304	0.87	0.13	2.5869	0.1131
0.37	0.63	0.7666	0.2331	0.88	0.12	2.7080	0.1056
0.38	0.62	0.7829	0.2356	0.89	0.11	2.8445	0.0979
0.39	0.61	0.7996	0.2379	0.90	0.10	3.0000	0.0900
0.40	0.60	0.8165	0.2400	0.91	0.09	3.1798	0.0819
0.41	0.59	0.8336	0.2419	0.92	0.08	3.3912	0.0736
0.42	0.58	0.8510	0.2436	0.93	0.07	3.6450	0.0651
0.43	0.57	0.8686	0.2451	0.94	0.06	3.9581	0.0564
0.44	0.56	0.8864	0.2464	0.95	0.05	4.3589	0.0475
0.45	0.55	0.9045	0.2475	0.96	0.04	4.8990	0.0384
0.46	0.54	0.9230	0.2484	0.97	0.03	5.6862	0.0291
0.47	0.53	0.9417	0.2491	0.98	0.02	7.0000	0.0196
0.48	0.52	0.9608	0.2496	0.99	0.01	9.9499	0.0099
0.49	0.51	0.9802	0.2499	1.00	0.00	infinito	0.0000
0.50	0.50	1.0000	0.2500				

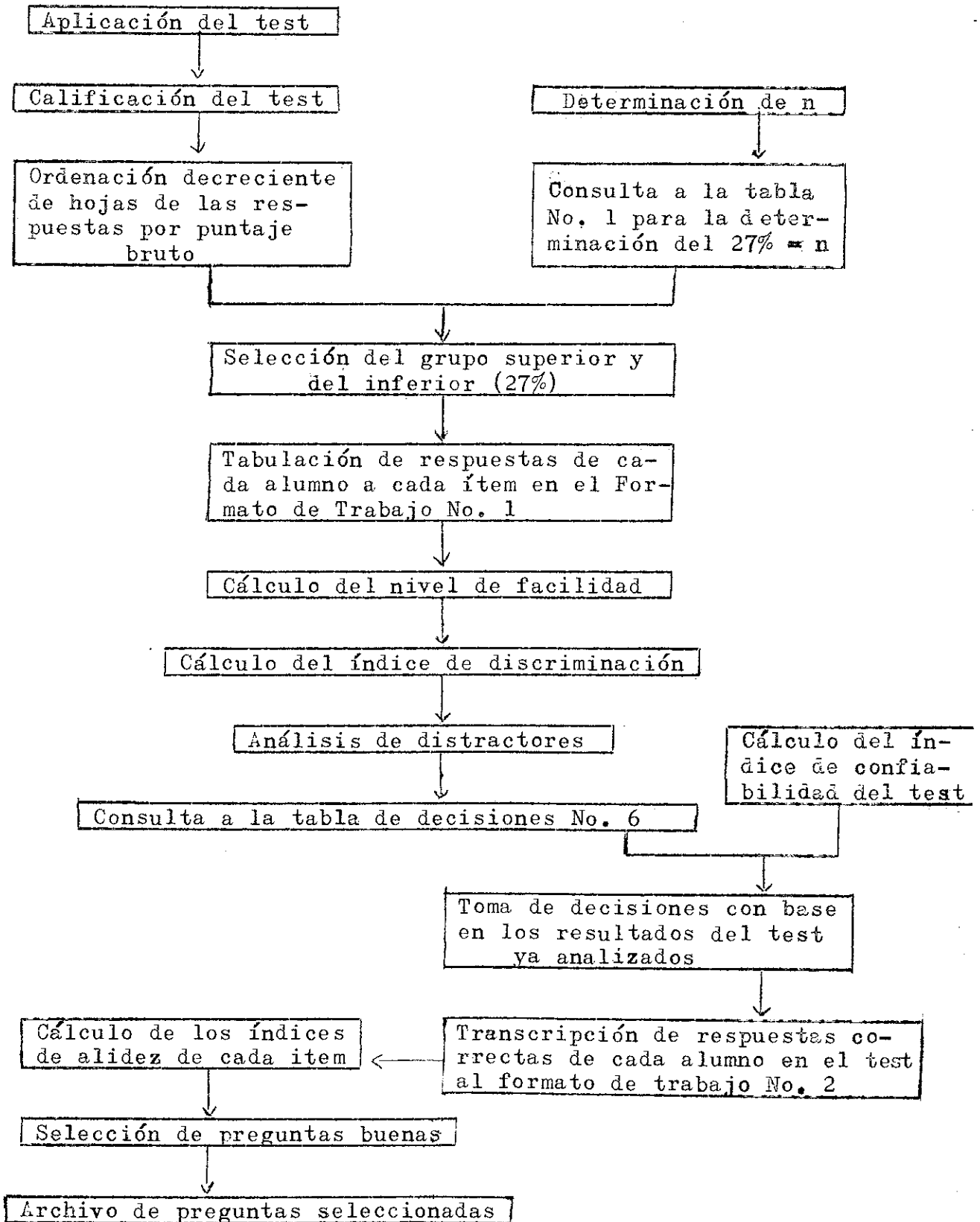
APÉNDICE E

Tabla No.4

K	$\frac{K}{K-1}$	K	$\frac{K}{K-1}$	K	$\frac{K}{K-1}$
20	1.0526	47	1.0217	74	1.0137
21	1.0500	48	1.0213	75	1.0135
22	1.0476	49	1.0208	76	1.0133
23	1.0455	50	1.0204	77	1.0132
24	1.0435	51	1.0200	78	1.0130
25	1.0417	52	1.0196	79	1.0128
26	1.0400	53	1.0192	80	1.0127
27	1.0385	54	1.0189	81	1.0125
28	1.0370	55	1.0185	82	1.0123
29	1.0357	56	1.0182	83	1.0122
30	1.0345	57	1.0179	84	1.0120
31	1.0333	58	1.0175	85	1.0119
32	1.0323	59	1.0172	86	1.0117
33	1.0313	60	1.0169	87	1.0116
34	1.0303	61	1.0167	88	1.0115
35	1.0294	62	1.0164	89	1.0113
36	1.0286	63	1.0161	90	1.0112
37	1.0278	64	1.0159	91	1.0111
38	1.0270	65	1.0156	92	1.0110
39	1.0263	66	1.0154	93	1.0109
40	1.0256	67	1.0152	94	1.0108
41	1.0250	68	1.0149	95	1.0106
42	1.0244	69	1.0147	96	1.0105
43	1.0238	70	1.0145	97	1.0104
44	1.0233	71	1.0143	98	1.0103
45	1.0227	72	1.0141	99	1.0102
46	1.0222	73	1.0139	100	1.0101

APÉNDICE F

DIAGRAMA (EN SECUENCIA) DE OPERACIONES PARA LA APLICACION DEL MODELO



APÉNDICE G

Tabla No.6

TABLA DE DECISIONES E INTERPRETACIÓN DE RESULTADOS

ÍNDICE DE DISCRIMINACION ($D = \frac{S-I}{n}$)

Valor de D	Clave	Descripción
0.40 o más	AD	Alta discriminación
0.30 - 0.39	RD	Regular discriminación
0.20 - 0.29	BD	Baja discriminación
0.10 o menos	ND	No discrimina

Si D resulta negativo, elimine el ítem.

ÍNDICE DE FACILIDAD ($F = \frac{S+I}{2n}$)

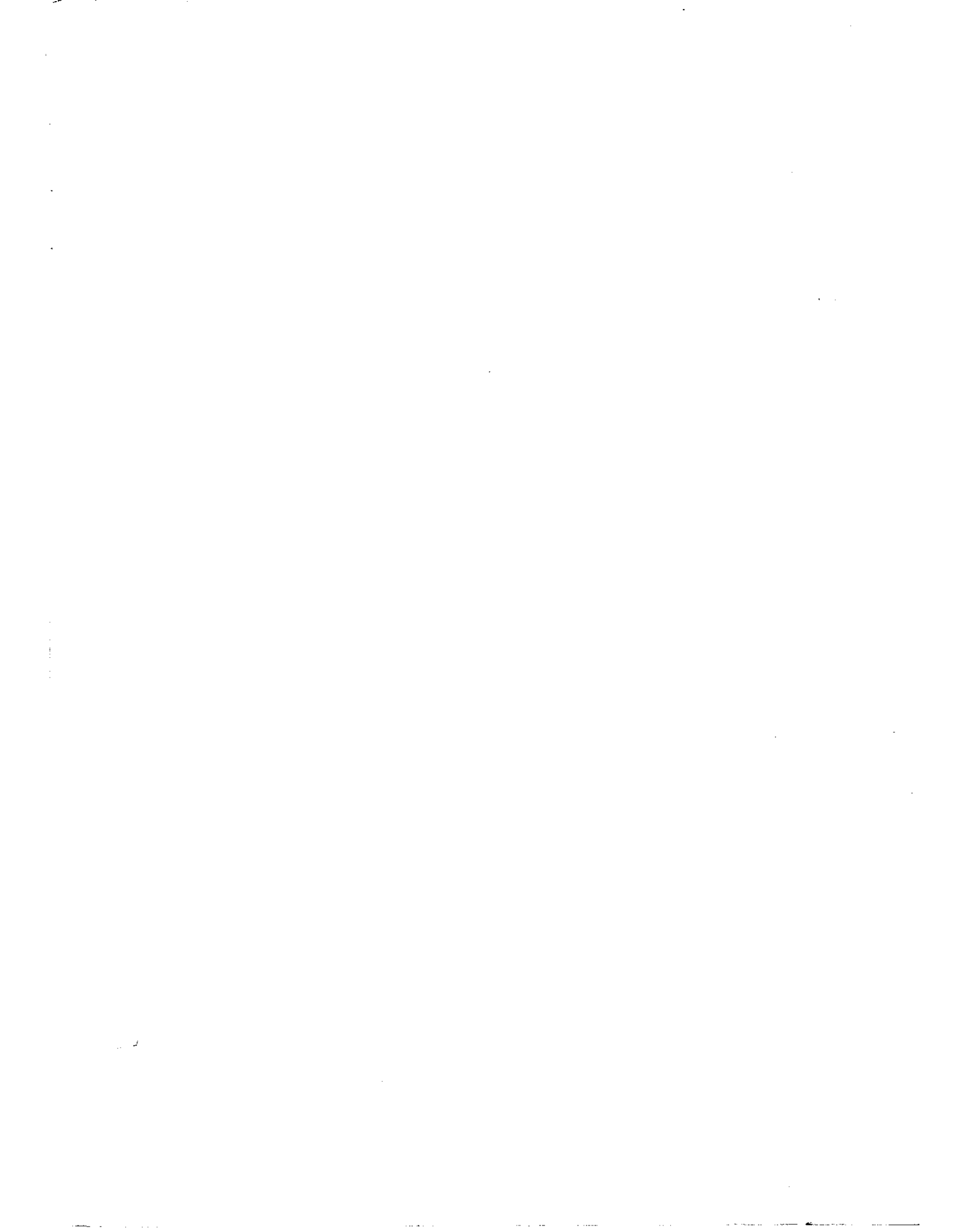
Valor de F	Clave	Descripción
0.76 - 1.00	F	Fácil
0.25 - 0.75	M	Medio
0.00 - 0.24	D	Difícil

Para cada una de las alternativas de respuesta de un ítem, verifique si S es mayor que I (el número de aciertos del grupo superior es mayor que el del inferior), de ser así, revise la alternativa de cuestión. Si esta situación se presenta en dos o más casos, elimine el ítem, pues será un indicio de que el distractor no funciona correctamente porque atrae más alumnos del grupo alto que del bajo.

Cuando aparecen varias omisiones en un ítem (señaladas con asterisco en el formato de trabajo No.1), puede deberse a que el ítem está mal estructurado conceptual o técnicamente, o a que es muy difícil. En este caso el procedimiento a seguir es revisar tanto el enunciado como las alternativas. Identifique el defecto y corrija revisando los procedimientos técnicos de estructuración de ítems. Posteriormente, pruebe los ítems que vayan sido reestructurados.

ÍNDICE DE CONFIABILIDAD DEL TEST (r)

Valor de R	Interpretación
0.80 o más	Buena confiabilidad (los resultados pueden usarse con confianza)
0.70 - 0.79	Confiabilidad baja (use los resultados con reservas)
0.70 o menos	Inconfiable (los resultados no son confiables para tomar decisiones)



FORMATO DE TRABAJO N:2

L_{Tg} $e_{st} \cdot d_{st} \cdot a_{n \cdot t_g}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	PB	PB-MT	PB-MT (PB-M)
1																							
2																							
3																							
4																							
5																							
6																							
7																							
8																							
9																							
10																							
11																							
12																							
13																							
14																							
15																							
16																							
17																							
18																							
19																							
20																							
SUMAS																							
$PI = \frac{S}{2n}$																							
MI																							
DM																							
$\frac{DM}{ST}$																							
$r_{Pois} = \frac{DM}{ST} \cdot \frac{PI}{V \cdot CI}$																							

← PBT
→ MT
← PBT

→ ST²
← PBT

→ ST
← PBT