

A mis padres

A la familia Ureña Farache

TE
UNIVERSIDAD DEL VALLE DE GUATEMALA
BIBLIOTECA

VALIDEZ DE "CONSTRUCTO" DE LOS ÍTEMES DE
UN TEST DE HABILIDAD MATEMÁTICA

por

RAMÓN DEL CARMEN BÁEZ UREÑA

TRABAJO DE INVESTIGACIÓN

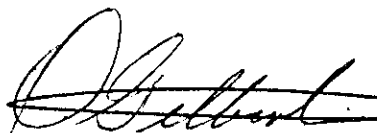
Presentado a la Facultad de Educación de la
Universidad del Valle de Guatemala
como requisito previo a
optar el Grado de
MAESTRÍA
en
MEDICIÓN, EVALUACIÓN E INVESTIGACIÓN EDUCATIVAS

UNIVERSIDAD DEL VALLE DE GUATEMALA

Guatemala, noviembre, 1974

VALIDEZ DE "CONSTRUCTO" DE LOS ÍTEMES DE
UN TEST DE HABILIDAD MATEMÁTICA

APROBADA:



Dr. Otto Gilbert

ÍNDICE

Capítulo		Página
I	INTRODUCCIÓN	1
II	INVESTIGACIONES RELACIONADAS	9
III	METODOLOGÍA	13
IV	RESULTADOS DE LA INVESTIGACIÓN	23
V	DISCUSIÓN	27
	REFERENCIAS BIBLIOGRÁFICAS	30

CAPÍTULO I

INTRODUCCIÓN

Un grupo de estudiantes del Programa de Maestría en Medición, Evaluación e Investigación Educativas que se desarrolló en la Universidad del Valle de Guatemala en 1974, tuvo a su cargo la elaboración de una prueba de habilidad matemática de nivel pre-universitario como parte de las actividades requeridas en el curso de "Medición y Construcción de Pruebas Educativas". Luego de elaborada esta prueba, se aplicó a grupos de estudiantes del nivel secundario; el objetivo inmediato de tal aplicación era hacer un análisis de ítemes para determinar su grado de dificultad así como su poder de discriminación.

Exposición de la naturaleza del tema

La construcción de un test de habilidad está basada en una serie de criterios distintos a los que se siguen al elaborar un test, por ejemplo, de rendimiento. En particular, los ítemes con los cuales se procura medir habilidad están elaborados en base a objetivos diferentes que aquéllos con los que se busca medir conocimientos. Sobre las características que debe reunir un ítem de un test de habilidad, entre otras cosas, se ha dicho lo siguiente:

El tipo de ítemes que compone estos test trata de explorar el ingenio del alumno al responderlo, antes que la mecánica del proceso para resolver un problema. Los problemas que figuran en un test de habilidad matemática deben corresponder a un nivel general de conocimiento inferior al que mediría un test de rendimiento para ese mismo grupo, y que se supone pudo haberse adquirido por experiencia, aún sin enseñanza sistemática.^{1/}

Estos criterios de construcción se tomaron en cuenta al elaborar el test de habilidad matemática pre-universitario. Sin embargo, cabría preguntar: ¿hasta qué punto se tiene la seguridad de que los problemas de este test corresponden a un nivel general de conocimiento que pudo haberse adquirido por experiencia, aún sin enseñanza sistemática? La respuesta es que no se tiene seguridad alguna sobre si en realidad se logró controlar este aspecto en cada ítem en particular. En otras palabras, no se tiene un grado de certeza tal que permita garantizar la validez de cada ítem del test en cuanto a la característica que debe medir; es decir, garantizar que el ítem mida habilidad matemática. Ante ese hecho, se hace necesario realizar un estudio sobre el test de habilidad matemática que ha sido construido con el fin de determinar la validez de sus ítemes.

1/ Cómo construir tests de habilidad matemática y de matemática (Guatemala: folleto mimeografiado, elaborado por la Universidad del Valle de Guatemala, 1973), pág. 2.

Un estudio de esta naturaleza conduce al establecimiento de la validez de "constructo" o validez de "construcción" del test de habilidad matemática de nivel pre-universitario. El objetivo de esta investigación será recopilar evidencia empírica para decidir la validez de los ítemes --y con ello la validez de "constructo"-- del test de habilidad matemática de nivel pre-universitario construido por los estudiantes del Programa de Maestría.

El camino a seguir será el siguiente: se clasificarán dos grupos de estudiantes de acuerdo a su habilidad numérica medida con el Test de Habilidad Numérica de los Test de Aptitud Diferencial y se hipotetizará que el grado de dificultad de los ítemes del test de habilidad matemática es igual en ambos grupos de habilidad, o lo que es lo mismo, que los ítemes no discriminan entre los estudiantes de alta y baja habilidad numérica.

Importancia del tema

En este apartado, los conceptos principales a destacar conciernen a la validez de "constructo", pero antes es importante señalar algunos aspectos sobre los tests de aptitud.

La construcción y uso de tests para detectar en los estudiantes aptitudes o habilidades ya sean generales o específicas,

reviste una importancia capital en el ámbito educativo. Al respecto, señala Lindeman

La medición de las aptitudes permite formular predicciones con respecto a lo que el alumno podría aprender si se le brindan las experiencias de aprendizaje apropiadas; también indica el nivel de aprendizaje o rendimiento que debe esperarse del alumno teniendo en cuenta sus aptitudes generales y específicas.^{2/}

El conocimiento de las capacidades potenciales de los alumnos es un factor decisivo en la enseñanza, sobre todo si se toman en cuenta las diferencias individuales. En los niveles de la escuela elemental es importante distribuir a los alumnos en subgrupos de aprendizaje de modo que reciban una enseñanza coherente con su aprendizaje previo y conforme con sus propias habilidades o aptitudes. En la escuela secundaria, el rol de los tests de aptitud tiene que ver con la orientación profesional y vocacional, señalándole a los futuros estudiantes de niveles de estudio más avanzados, las probabilidades de éxito en determinadas áreas del saber.

Por lo general, la medición de las aptitudes se lleva a cabo mediante el uso de tests estandarizados. "El término

2/ Richard H. Lindeman. Tratado de Medición Educativa (Buenos Aires: Editorial Paidós, 1971), pág. 16.

estandarizados se refiere al hecho de que estos tests están contruidos para ser administrados, evaluados e interpretados en condiciones típicas".^{3/} La etapa última en la preparación de un test es su estandarización; el proceso de estandarización conlleva también un proceso de validación del test, el cual supone cinco tipos de validez. Son ellos: validez aparente, validez de contenido, validez predictiva, validez concurrente y validez de "constructo". Los comentarios siguientes concier- nen a la validez de "constructo" que es el objetivo de este es- te estudio.

Validez de "constructo"

El concepto de validez de "constructo" fue creado en 1954 por el comité técnico designado por la Asociación Americana de Psicología (American Psychological Association). El propósito de este comité fue definir la terminología y formular reglas para la estandarización de tests.^{4/}

La validez de "constructo" se refiere a la medida en que un test proporciona información acerca de una característica significativa del individuo^{5/}. En el contexto de esta

3/ Lindeman, op.cit., pág. 17.

4/ David Magnusson. Teoría de los tests (México: Editorial Trillas, 1972), pág. 160

5/ Lindeman, op.cit., pág. 53.

investigación, la característica significativa o "constructo", es la habilidad matemática. Refiriéndose específicamente a los ítemes, Lindeman ha señalado:

Para lograr que un test tenga una alta validez de "constructo", el maestro debe otorgar a los ítemes de aquél una calidad tan elevada como sea posible y lograr un equilibrio adecuado entre ítemes fáciles y difíciles.^{6/}

Este tipo de validez es útil e importante para toda clase de tests psicológicos, especialmente aquellos que miden rasgos para los cuales no hay un criterio externo; por ejemplo, los tests de personalidad. Los estudios de validez de "constructo" constituyen el único medio de definir con más claridad justamente lo que está siendo medido por una prueba específica. La importancia de la validez de "constructo" es puesta más claramente de manifiesto en los siguientes párrafos:

Los tests muy raramente miden con exactitud lo que se halla implicado en su denominación. A pesar de eso, no se puede interpretar un test hasta tanto se conozca cuáles son los factores que determinan las puntuaciones. La determinación de la validez de "constructo" conlleva a un análisis de la significación de las puntuaciones del test expresado en conceptos psicológicos.^{7/}

6/ Ibid

7/ Lee J. Cronbach. Fundamentos de la exploración psicológica (Madrid: Editorial Biblioteca Nueva, 1972), pág. 165.

Declaración de la hipótesis de investigación

Los ítemes del test de habilidad matemática discriminan entre los estudiantes de alta y baja habilidad numérica, al nivel de significación 0.05.

Identificación de variables

Las variables de este estudio son:

A. Una variable predictora o variable independiente determinada por los puntajes brutos del test de Habilidad Numérica de los Tests de Aptitud Diferencial. Esta variable tiene dos niveles:

1. Nivel de habilidad alto

Está determinada por una muestra que representa el 27 por ciento de los estudiantes que obtuvieron mayor puntajes de la población total examinada.

2. Nivel de habilidad bajo

Está determinada por el 27 por ciento que obtuvieron los puntajes más bajos en la población total examinada.

B. Una variable predicha o variable dependiente constituida por los índices de discriminación de los ítemes del test de habilidad matemática encontrados para los dos grupos de habilidad diferente.

Limitaciones de este estudio

Los resultados de esta investigación no se pueden extender más allá del propio test bajo estudio, dado que se trata de un instrumento no estandarizado, del cual no se conoce característica alguna, salvo la que se encuentre en el presente estudio. Tampoco podrá generalizarse más allá de la población con la que se hará el estudio, es decir, los estudiantes de cuarto y quinto grado del bachillerato del Colegio Americano de Guatemala y los estudiantes de cuarto semestre de la Escuela Politécnica de Guatemala, debido a que dicha población no es representativa de la población guatemalteca de ese nivel.

CAPÍTULO II

INVESTIGACIONES RELACIONADAS

En las primeras décadas de este siglo, cuando cobró auge la construcción de tests para detectar características significativas en el individuo, la forma de llegar a determinar la validez de "constructo" de tales instrumentos era poco objetiva. En esos momentos, la validez de "constructo" de un test la decidía la opinión de un experto. Binet, por ejemplo, en sus primeros trabajos --y basado en sus propios juicios-- elaboró una serie de tareas con las cuales buscaba distinguir entre estudiantes normales y retrasados. Muy pronto, sin embargo, se puso de manifiesto que no bastaban los juicios de los expertos cuando se trataba de encontrar la relación existente entre una característica determinada del individuo y los tipos de comportamientos asociados con ella. Esto se hizo evidente en uno de los primeros tests de Binet al descubrirse que entre los individuos que lograban repetir satisfactoriamente una cantidad determinada de dígitos enunciados por el examinador, había algunos que se hallaban significativamente por debajo de lo normal en lo que respecta a la habilidad mental general y al rendimiento en muchas tareas de la escuela.^{8/} Este hecho obligó

8/ Lindeman, op.cit., pág. 53.

en lo adelante, a buscar nuevas formas de determinar la validez de "constructo", basándose sobre todo en recursos estadísticos. Los trabajos posteriores sobre tests educacionales y psicológicos siguieron por esta vía, estableciéndose la validez de "constructo" mediante la demostración estadística del grado de relación entre una característica determinada del individuo y los tipos de comportamientos asociados con ella. Las técnicas y procedimientos estadísticos usados con mayor frecuencia son las correlaciones con otros tests, el análisis factorial y los efectos inducidos experimentalmente. Para todos los tests estandarizados que se conocen se ha establecido la validez de "constructo" por uno o más de estos métodos estadísticos. A continuación se detallan estudios sobre el establecimiento de la validez de "constructo" de la Prueba de Habilidad Mental Otis-Lennon.

Las Pruebas Otis-Lennon fueron diseñadas para evaluar la habilidad mental general. En el desarrollo de estas pruebas se definió la habilidad mental general como un compuesto de las habilidades verbales-educacionales que deben ser usadas en las diversas situaciones que requieren razonamiento abstracto con materiales verbales, numéricos y de figuras.

Las Pruebas Otis-Lennon se correlacionaron con otras medidas dentro del dominio general de habilidad mental-aptitud

entre las cuales está la Escala de Inteligencia Stanford-Binet y las Matrices Progresivas Standard de Raven. Usando una muestra de 256 sujetos, el Nivel Elemental I de Otis-Lennon dado al final del grado 2 y el Stanford-Binet dado casi 2 años antes, correlacionaron 0.60. Con las Matrices de Raven, usando muestras de 284, 288 y 244, se obtuvieron correlaciones en los grados 5, 7 y 11 que se agrupan en 0.60; este hecho se entiende como una relación substancial entre las dos medidas.

Se correlacionó también el Nivel Primario II de Otis-Lennon y las varias subpruebas de las Pruebas de Rapidez Metropolitan (Metropolitan Readiness Tests) y el Análisis de Rapidez de Lectura del Murphy-Durrell. Las correlaciones totales obtenidas para Otis-Lennon de 0.65 (Metropolitan Readiness Tests) y 0.65 (Murphy-Durrell) fueron del nivel esperado. Las muestras de sujetos para estos estudios fueron 215 y 71 sujetos.

Se hizo un estudio especial diseñado para determinar la equivalencia entre las pruebas Otis-Lennon y la Serie Otis de Puntajes Rápido con muestras que comprendieron desde 672 sujetos a muestras de 1,289 estudiantes. Se encontraron correlaciones variando de 0.70 a 0.91 con una mediana de 0.82 para los grupos estudiados de grado único. Los resultados obtenidos indican un grado muy alto de traslape entre las dos series, lo cual se esperaba, debido a que las pruebas Otis-Lennon representan una edición revisada de la Serie de Puntajes Rápido.

Se han hecho correlaciones entre las Pruebas Otis-Lennon y "Cooperative School and College Ability Tests" (SCAT). También con las Pruebas Ohio Survey: Sección de Aptitud. Las correlaciones con las pruebas SCAT y la Ohio Survey varían desde 0.70 a 0.90 para muestras de sujetos que varían entre 67 y 514. Esto es un indicio de que estas medidas y las Otis-Lennon están evaluando habilidades similares.

En el proceso de establecimiento de la validez de "constructo" de la Prueba Otis-Lennon, los resultados de dicha prueba fueron correlacionados con resultados de los Tests de Aptitud Diferencial y se encontró un coeficiente de correlación de 0.76 para Razonamiento Verbal, 0.74 para Habilidad Numérica, 0.65 para Razonamiento Abstracto y 0.85 para Verbal más Numérico. La muestra examinada fue de 122 sujetos.

Resumiendo, en la mayoría de las correlaciones que se han hecho entre los resultados de las Pruebas Otis-Lennon y otras medidas dentro del dominio general de habilidad mental-aptitud, se han encontrado resultados notables que indican la similitud en lo que están midiendo.

CAPÍTULO III

METODOLOGÍA

Tipo de investigación

Esta investigación pertenece al tipo llamado comúnmente investigación "aplicada" o investigación no experimental. Este tipo de investigación procura recopilar, organizar e interpretar datos para la solución de problemas inmediatos. En consecuencia, este estudio empírico descansa en un diseño de investigación no experimental en el que los sujetos en base a los cuales se realizó, no fueron sometidos a un tratamiento experimental. En esta investigación no hubo una asignación al azar de los sujetos que participaron en el estudio; sin embargo, las variables fueron manipuladas por el investigador, previa a la realización del estudio.

Muestra

La muestra utilizada para la obtención de los datos con los cuales se pueda decidir entre no aceptar o no rechazar las hipótesis de investigación, consistió en 76 estudiantes seleccionados de una población de 139 sujetos. Esta población la componen los estudiantes de cuarto y quinto grado de bachillerato del Colegio Americano de Guatemala y los estudiantes del cuarto semestre de la Escuela Politécnica de la misma ciudad.

Estos últimos tienen un nivel académico equivalente al quinto grado del bachillerato guatemalteco. Es importante aclarar que los ítemes 18 y 19 fueron analizados con una muestra diferente dado que en un principio fueron mimeografiados con errores; por tal motivo, se eliminaron los resultados de la primera aplicación del test en lo que respecta a los dos ítemes mencionados. En la segunda aplicación del test, los ítemes 18 y 19 habían sido corregidos. La forma para seleccionar los 76 estudiantes fue la siguiente: se aplicó el test de Habilidad Numérica (Forma B) de los Tests de Aptitud Diferencial a los 139 estudiantes; luego de calificar a mano las hojas de respuesta, se ordenaron de acuerdo al puntaje bruto obtenido, de mayor a menor. Se tomó entonces el 27 por ciento de los puntajes más bajos y el 27 por ciento de los puntajes más altos y el resultado fue 38 sujetos para cada grupo, es decir, 76 en total.

Cuando se hacen análisis de ítemes con grandes muestras de sujetos, se aplica el criterio del 27 por ciento, es decir, tomar el 27 por ciento de puntajes bajos y el 27 por ciento de los altos. Esta es una forma --entre otras-- para el análisis de ítemes; la muestra así obtenida resulta representativa de la población entera. Cuando la población es pequeña se usa el criterio del 50 por ciento, o sea, un 50 por ciento bajo y un 50 por ciento alto.

Para el presente estudio la población era de 139 estudiantes, sin embargo, se usó el criterio del 27 por ciento. Esto se hizo así para tener grupos bien diferenciados respecto de su habilidad. Se dejó fuera de estudio al 46 por ciento de los estudiantes, los cuales obtuvieron puntuaciones promedio comprendidas entre el grupo bajo y el grupo alto. Los grupos seleccionados para este estudio diferían en por lo menos 10 puntos respecto de su habilidad; o sea, el puntaje más bajo del grupo alto fue 28 en el test de Habilidad Numérica, mientras que el puntaje más alto del grupo bajo en dicho test fue 18. Para los ítemes 18 y 19 del test de habilidad matemática se hizo un análisis aparte. El ítem 18 se analizó utilizando una población de 60 sujetos; 64 estudiantes tomaron el test de habilidad matemática con el ítem 18 corregido, pero sólo 60 de ellos tomaron el test de Habilidad Numérica en base al cual se establecieron los grupos de diferente habilidad. En este caso se aplicó el criterio del 50 por ciento bajo y el 50 por ciento alto, o sea, 30 estudiantes en cada grupo. El ítem 19 se analizó utilizando una población de 104 sujetos que tomaron el test de habilidad matemática con dicho ítem corregido. En este caso se aplicó el criterio del 27 por ciento, resultando 28 estudiantes para el grupo bajo y 28 para el grupo alto; es decir, 56 estudiantes en total.

Instrumentos usados para la obtención de datos

Para llevar a cabo esta investigación se hizo uso de dos tests: el test de Habilidad Numérica (Forma B) de los Tests de Aptitud Diferencial y el test de habilidad matemática objeto de este estudio. En base al primero se establecieron los dos grupos de habilidad, altos y bajos; en base al segundo se hizo el análisis de los ítemes con los grupos previamente seleccionados con el de Habilidad Numérica.

Los Tests de Aptitud Diferencial constituyen una batería de tests estandarizados para los Estados Unidos de Norteamérica. Después de su estandarización allí han sido traducidos a otros idiomas y adaptados y normalizados en otros países; es decir, se han hecho cambios a la versión original y se han derivado normas de interpretación propias de algunos países en donde se usa.

El Colegio Americano de Guatemala posee los derechos de publicación sobre estos Tests; los mismos son ampliamente usados por la Institución y actualmente se lleva a cabo su estandarización para Guatemala. Se han hecho adaptaciones y se poseen normas guatemaltecas para interpretar sus resultados. Debido a estas razones, los puntajes obtenidos en el test de Habilidad Numérica se aceptarán como confiables para los propósitos de este estudio.

El test de Habilidad Numérica tiene 40 ítemes de selección múltiple. Una respuesta correcta se valora en uno y una respuesta incorrecta se valora en cero; así, un estudiante podría obtener 40 de punteo bruto (puntaje máximo) si todas sus respuestas resultan correctas. Para determinar el punteo bruto obtenido por un estudiante en el test se aplica una fórmula de corrección que procura eliminar el posible factor adivinación al responder a un ítem. La fórmula es

$$C - 1/4^I$$

que quiere decir: número de respuestas correctas menos un cuarto de las incorrectas. Si un estudiante responde a 32 ítemes del test y entre sus respuestas 24 son correctas y 8 son incorrectas, su punteo bruto será:

$$\begin{aligned} \text{punteo bruto} &= C - 1/4^I \\ &= 24 - (1/4 \times 8) \\ &= 24 - 2 = 22 \end{aligned}$$

En este estudio se consideró sólo el punteo bruto. Teóricamente, un punteo bruto podría ser negativo, o sea, menor que cero; en tales casos se consideraría igual que cero.

En cuanto a las características de los Tests de Aptitud Diferencial, sus autores han expresado lo siguiente:

Los Tests de Aptitud Diferencial constituyen un instrumento completo y científico para la medición de las habilidades de los estudiantes de las escuelas secundarias y de las universidades. Fueron elaborados para servir propósitos de selección y de orientación educativa y vocacional. Se usan también en la selección de empleados.^{9/}

Sobre el test de Habilidad Numérica en particular, sus autores han expresado lo siguiente:

El test de Habilidad Numérica mide la habilidad para razonar con números, para manipular relaciones numéricas y para operar inteligentemente con materiales cuantitativos. Evalúa, asimismo, la comprensión de las relaciones numéricas y la facilidad para manejar conceptos numéricos. También predice el éxito en matemáticas, física, química e ingeniería; así como en trabajo de laboratorio, contaduría, manufactura de instrumentos y en otras ocupaciones y profesiones relacionadas con las ciencias físicas; es decir, en todas aquellas para las cuales el pensamiento matemático es esencial.^{10/}

Estas son, en resumen, las características de los Tests de Aptitud Diferencial y en particular el test de Habilidad Numérica. Ahora es necesario describir el test de habilidad matemática objeto de esta investigación. Sobre este test no pueden adelantarse muchas características puesto que es el que

9/ George K. Bennet; Harold G. Seashore; Alexander G. Wesman. Manual de instrucciones de los Tests de Aptitud Diferencial, Formas A y B. (Editado por el Colegio Americano de Guatemala, 1970), pág. 1.

10/ Ibid., pág. 6.

está en estudio. Al elaborarse se tomaron en cuenta los criterios de validez aparente así como los de validez de contenido; su validez de "constructo" está en estudio y su validez predictiva deberá ser objeto de futuras investigaciones.

La validez aparente y la de contenido no se determinan experimentalmente; más bien descansan en el juicio de expertos en la materia. Un test que parece a simple vista bueno para un propósito determinado, tiene lo que se llama "validez de fachada" o "validez aparente".^{11/} Por otra parte, Adams señala lo siguiente:

La validez aparente concierne a la medida en que un test parece medir la información y las capacidades relevantes. La única forma de apreciar la validez aparente, así como la idoneidad del nivel de dificultad del test, consiste en estudiar directamente las preguntas del test.^{12/}

Los juicios de validez aparente llevan a decidir si los ítemes de un test son apropiados en la situación específica en que han de usarse; por tanto, la validez aparente debe ser el primer criterio para incluir un ítem en un test. Básicamente se necesita que los ítemes estén elaborados de acuerdo al grado

11/ Cronbach, op.cit., pág. 211

12/ Georgia Sachs Adams. Medición y evaluación en educación, psicología y "Guidance" (Barcelona: Editorial Herder, 1970), pág. 211.

de madurez y nivel académico de aquellos que toman el test. Esta característica se da como lograda en el test de habilidad matemática.

Para la validez de contenido se elaboró una tabla de especificaciones de doble entrada, es decir, en la que se incluyen objetivos y contenidos didácticos. Conforme a esta tabla se elaboró el test (ver Cuadro 1). Para una breve explicación de la tabla obsérvese, por ejemplo, la casilla donde aparece el número 5; el número 5 significa que hay cinco ítemes de aritmética del tipo juicio numérico. El número 8 que aparece en otra casilla significa que hay en total 8 ítemes de cómputo, de los cuales 3 son sobre teoría de conjuntos, 2 de aritmética y 3 de álgebra. Al mismo tiempo, 8 equivale al 32 por ciento del porcentaje total de ítemes.

CUADRO 1

Tabla de especificación para la construcción del test de habilidad matemática

Contenido	Proceso			Totales	Porcentaje
	Cómputo	Jucio numérico	Pensamiento de relación		
Teoría de conjunto	3	2	1	6	24
Aritmética	2	5	-	7	28
Algebra	3	3	-	6	24
Geometría	-	-	6	6	24
Totales	8	10	7	25	
Porcentaje	32	40	28		100%

En la categoría de cómputo se incluyen ítemes cuya respuesta depende del resultado exacto de una manipulación rutinaria, que debió haber estado sujeta a ejercitación en el aula. En la categoría juicio numérico se incluyen ítemes cuyo punto principal es el de reconocer ciertas características de los números que se presentan, y/o aplicar estas características en la solución de un problema específico. Por último, en la categoría pensamiento de relación se incluyen ítemes que requieren aplicar o hacer inferencias de relaciones dadas entre variables u objetos.^{13/}

Tratamiento estadístico

El tratamiento estadístico abarca tres aspectos principales:

- A. Determinación del coeficiente de confiabilidad del test de habilidad matemática. Por la fórmula $K-R_{20}$ o fórmula de Kuder-Richarson se determinó el coeficiente de confiabilidad del test de habilidad matemática. A partir del coeficiente muestral encontrado " r_{tt} " se hipotetizó que el coeficiente poblacional " ρ " (rho) es igual a cero para un nivel de significación estadística de 0.05; o sea, se estableció que la hipótesis

13/ Cómo construir tests de habilidad matemática y de matemática (Folleto mimeografiado, elaborado por la Universidad del Valle de Guatemala, 1973), pág. 4.

nula era $H_0: \rho = 0$ y mediante la prueba "t" de "Student" se ensayó la hipótesis H_0 .

B. Determinación del grado de dificultad de los ítems del test de habilidad matemática para los dos grupos de habilidad.

C. Uso del estadístico X^2 . Se usó el estadístico X^2 (chi cuadrado) para determinar si hay diferencias estadísticamente significativas entre el grado de dificultad de los ítems en el grupo de habilidad numérica alto y el de habilidad numérica bajo. Cuando los cálculos de X^2 se hacen con un grado de libertad, se usa una fórmula corregida.^{14/} En este caso se usó la fórmula de X^2 corregida.

El estadístico X^2 contrasta las frecuencias esperadas y las frecuencias observadas en la ocurrencia de una serie de sucesos. En el contexto de esta investigación, las frecuencias observadas son el grado de dificultad del grupo de habilidad alto y el grado de dificultad del grupo de habilidad bajo para cada ítem. Las frecuencias esperadas se consideraron como el valor promedio de las frecuencias observadas.

14/ Murray R. Spiegel. Estadística (México: Editorial Libros McGraw-Hill de México, 1970), pág. 203.

CAPÍTULO IV

RESULTADOS DE LA INVESTIGACIÓN

Se computó para el test de habilidad matemática un coeficiente de confiabilidad $K-R_{20}$ de Kuder-Richarson usando una muestra de 64 sujetos; este número de estudiantes tomó el test en su forma corregida. Los puntajes obtenidos por los 64 estudiantes aparecen en el Cuadro 2.

CUADRO 2

Frecuencias de los puntajes de los sujetos con los cuales se determinó el coeficiente de confiabilidad (n=64)

Puntajes	Frecuencia	Puntajes	Frecuencia
5	6	15	6
7	3	16	6
8	5	17	1
9	3	18	3
10	4	19	1
11	4	20	1
12	3	21	1
13	8	24	1
14	8		

La varianza de los puntajes que aparecen en el Cuadro 2 fue $S^2 = 17.95$. El coeficiente de confiabilidad encontrado fue

$r_{tt}=0.76$ (véase Cuadro 3). El valor encontrado para el estadístico "t" de "Student" fue $t=14.17$.

CUADRO 3

Valores del producto pq para obtener el coeficiente de confiabilidad, en donde p = proporción de estudiantes que responden correctamente un ítem, y q = proporción de estudiantes que responden incorrectamente (n=64)

Ítem	pq	Ítem	pq
1	.24	14	.25
2	.06	15	.25
3	.17	16	.24
4	.00	17	.25
5	.20	18	.25
6	.20	19	.24
7	.22	20	.21
8	.11	21	.20
9	.23	22	.11
10	.25	23	.04
11	.25	24	.17
12	.24	25	.24
13	.21		

Respecto del coeficiente de confiabilidad, se ensayó la hipótesis $H_0: \rho=0$. Según un ensayo unilateral de una distribución de "Student" al nivel de significación 0.05, se rechazará H_0 si

$t > t_{.95} = 1.67$ para $(64-2) = 62$ grados de libertad. Como $t = 14.17$ fue mayor que $t_{.95} = 1.67$, se rechazó la hipótesis nula; en otras palabras, no se acepta que el coeficiente de confiabilidad del test para una población determinada sea igual a cero.

En el Cuadro 4 se muestran los índices de dificultad de cada ítem del test, correspondientes a los dos grupos de habilidad --grupo alto en habilidad numérica y grupo bajo--, así como el valor de X^2 (chi cuadrado) que da el índice de discriminación de cada ítem con respecto a los grupos de habilidad. La hipótesis en esta investigación es que al nivel de significación 0.05 los ítems del test de habilidad matemática discriminan entre estudiantes de alta y baja habilidad numérica. Por tanto, los valores de X^2 fueron contrastados con $X^2_{.95} = 3.84$ con un grado de libertad. En todos aquellos casos en que X^2 observado fue mayor que $X^2_{.95}$ se aceptó la hipótesis. Como puede observarse en el Cuadro 4, en 19 de los 25 ítems la hipótesis es aceptada; para los ítems 2, 4, 14, 20 y 25 la hipótesis no es aceptada. Cuadro 4 en la siguiente página.

CUADRO 4

Indices de dificultad de los ítemes del test de
habilidad matemática según los grupos de
habilidad numérica y el valor de X^2

Ítem	Índice de dificultad		X^2	$X^2_{0.95}$
	Grupo alto	Grupo bajo		
1	66	26	16.53	3.84
2	100	87	<u>0.77</u>	
3	89	37	20.64	
4	97	89	<u>0.26</u>	
5	45	24	5.80	
6	37	21	3.87	
7	74	47	5.59	
8	29	18	<u>2.13</u>	
9	61	16	25.14	
10	61	32	8.43	
11	63	34	8.08	
12	79	55	3.95	
13	53	11	26.26	
14	47	32	<u>2.48</u>	
15	68	37	8.57	
16	68	18	27.92	
17	55	29	7.44	
18	70	47	4.14	
19	57	25	11.72	
20	89	74	<u>1.20</u>	
21	50	11	23.67	
22	21	3	12.04	
23	13	0	11.08	
24	89	55	7.56	
25	50	63	<u>1.27</u>	

CAPÍTULO V

DISCUSIÓN

El primer hecho importante a destacar en los resultados de este estudio se refiere al coeficiente de confiabilidad encontrado para el test de habilidad matemática, el cual fue $r_{tt}=0.76$. La hipótesis de que el valor hallado no difería de cero para una población determinada fue ampliamente rechazada al hacerse la prueba de significado estadístico según un ensayo unilateral de una distribución de "Student" al nivel de significación de 0.05 ($t_{.95}=1.67$ contra $t=14.17$). Esto significa que el valor encontrado es una buena medida del coeficiente de confiabilidad. El uso de "t" se basó en el supuesto de que el estadístico que se está probando se distribuye normalmente. Esta suposición no fue estrictamente comprobada en este estudio, sin embargo, el tamaño grande de la muestra utilizada podría superar esa limitación.

En cuanto a los índices de dificultad puede observarse en el Cuadro 4 que en todos los ítemes --excepto el 25-- se mantiene una correspondencia en el sentido de que a mayor habilidad mayor es el índice de dificultad del ítem. En el grupo de habilidad alto el grado de dificultad para el ítem 1 fue 66, mientras que para el grupo bajo de habilidad el índice de dificultad

fue 26, y así sucesivamente. Es importante señalar que el grado de dificultad está expresando el porcentaje de estudiantes que responde correctamente el ítem, o sea, a mayor porcentaje de estudiantes que responde el ítem, menor es su grado de dificultad. En cuanto al ítem 25 hay que señalar que probablemente esté formulado ambigüamente, puesto que para el grupo de mayor habilidad el índice de dificultad fue menor.

Si para cada ítem, exceptuando el 25, se mantiene la diferencia en el grado de dificultad según se mantiene la diferencia en habilidad, o sea, a mayor habilidad menor dificultad o a menor habilidad mayor dificultad, entonces lógicamente los ítems del test de habilidad matemática deben estar realmente midiendo habilidad. En otras palabras, en el 96 por ciento de los casos los ítems probablemente están midiendo habilidad matemática. De no ser así, no se daría la correspondencia entre grado de dificultad y el nivel de habilidad.

Por otra parte, como la diferencia entre los dos valores de dificultad varía para cada ítem, hay que decidir cuáles de estas diferencias son estadísticamente significativas. De acuerdo con la prueba X^2 (chi cuadrado), de los 24 ítems en los cuales se mantiene la correspondencia entre grado de dificultad y nivel de habilidad, sólo 19 de ellos discriminan en forma estadísticamente significativa.

Conclusiones

Las conclusiones fundamentales de este estudio consignan que el valor encontrado para el coeficiente de confiabilidad del test de habilidad matemática de nivel pre-universitario es una medida fiel de su consistencia interna.

En cuanto a los ítemes del test de habilidad matemática se concluye que 19 de los 25 ítemes miden habilidad matemática; por tanto, el test de habilidad matemática se debe elaborar en base a estos 19 ítemes.

Recomendaciones para futuras investigaciones

Para futuros estudios podría estructurarse un test teniendo en cuenta sólo aquellos ítemes que resultaron consistentes con la hipótesis experimental, o sea, aquellos que discriminan en forma estadísticamente significativa entre estudiantes de habilidad alta y de habilidad baja. Sin embargo, debe hacerse previamente un trabajo para asegurar que todas las opciones de cada ítem funcionan adecuadamente. Luego de realizarse esta labor podría determinarse la validez predictiva del instrumento.

REFERENCIAS BIBLIOGRÁFICAS

- Adams, Georgia Sachs. Medición y evaluación en educación, psicología y "Guidance". Barcelona: Editorial Herder, 1970. 822 pp.
- Anastasi, Anne. Tests psicológicos. Madrid: Editorial Aguilar, 1973. 686 pp.
- Campbell, Donald T. y Stanley, Julian C. Diseños experimentales y cuasiexperimentales en la investigación social. Buenos Aires: Editorial Amorrortu editores, 1973. 164 pp.
- Cronbach, J. Lee. Fundamentos de la exploración psicológica. Madrid: Editorial Biblioteca Nueva, 1971. 834 pp.
- Dieppa, Jorge J. La investigación científica en la orientación profesional. (Trabajo leído ante el Primer Seminario Regional de Actualización en Desarrollo de Pruebas Educativas celebrado en la Universidad del Valle de Guatemala el día 13 de junio de 1974). 39 pp.
- Karmel, Louis J. Medición y evaluación escolar. México: Editorial Trillas, 1974. 552 pp.
- Lindeman, Richard H. Tratado de medición educacional. Buenos Aires: Editorial Paidós, 1971. 224 pp.
- Magnusson, David. Teoría de los tests. México: Editorial Trillas, 1972. 320 pp.

Spiegel, Murray R. Estadística. México: Editorial Libros
McGraw-Hill de México, 1970. 360 pp.

Wiersma, William. Research Methods in Education. Philadelphia:
J. B. Lippincott Company, 1969.

