

“ADAPTACION PSICOMETRICA DEL TEST DE OPERACIONES FORMALES COMBINATORIAS (T.O.F.C.) DE LONGEOT DE ACUERDO AL MODELO DE MOKKEN”

Ps. Luis Miguel Escurra Mayaute

Ps. Ana Delgado Vásquez

Ps. María Rosario Quesada Murillo

Ps. Gerardo Rivas Castro

Ps. Julio Santos Islas

Ps. Juan Pequeña Constantino

RESUMEN

Se evaluaron las características psicométricas del test de operaciones formales combinatorias de Longeot (T.O.F.C.), según el modelo de Guttman y el modelo de Mokken en una muestra de 260 estudiantes de la facultad de Psicología de la UNMSM.

El resultado del análisis del escalograma de Guttman indica que se obtiene un índice de reproductibilidad de 0.53. En tanto que el análisis de ítems en base al modelo de Mokken, permite observar que los coeficientes $H(i,j)$ de Loevinger y los coeficientes Delta Star (i,j) corresponden al modelo estudiado. El coeficiente H de Loevinger de la escala total asciende a 0.81, lo cual corrobora la adecuación del T.O.F.C. de Longeot al Modelo de Mokken. Adicionalmente se observa que la estimación de la confiabilidad rho es de 0.82

El análisis gráfico de las curvas características de los ítems permite observar que las curvas son paralelas y no se intersectan, verificándose el supuesto de Doble monotonía que debe existir entre los ítems del T.O.F.C.

El instrumento demostró poseer validez de construcción, el análisis factorial exploratorio en base al procedimiento de los componentes principales y la rotación Varimax, permitió demostrar la existencia de un solo factor. La confiabilidad por consistencia interna a través del coeficiente Kuder-Richardson 20, ascendió a 0.74. Los resultados generales indican que el 38.50% de la muestra corresponde al nivel Concreto, el 33.50% al nivel Formal A y el 28.00% llega al nivel Formal B. Las comparaciones por sexo y por promoción indican que no existen diferencias significativas entre los grupos.

ABSTRACT

Psychometrical properties were evaluated of combining formal operations of Longeot (T.O.F.C.), according to Guttman and Moken model between 260 students of psychology faculty of the UNMSM.

The result of Guttman's analysis indicates that an index of reproductiveness is obtained of 0,53. The analysis of items according to Mokken's model allows to notice that the $H(i,j)$ Loevinger and Delta Star (i,j) coefficients correspond to studied model. The Loevinger's H coefficients rises to 0,81, wich one corroborates Longeot's adjust ment of T.O.F.E. to Mokken's model the valuation of rho reliability is 0,82.

The graphic analysis of curves of items allows to notice that curves are parallel and they don't intersect, verifying the doble monotony that exist between items of T.O.F.C.

The instrument smowed having a validity of construction, factorial explorative analysis on the basis of procedure of main components and the Varimax rotation, allowed to smow the existence of only factor. The confiability by internal consistency through Kuder Richardson 20 coefficient, rised to 0,74. The general results indicate that 38,5% of sample correspond to concrete level, the 33,5% to formal a level, and the 28% achieves to Formal B level. The comparisons by sex and promotion indicate that not exist significant differences between the groups.

La medición psicológica ha variado en los últimos tiempos influenciada por los diferentes paradigmas de investigación, pues hasta hace décadas pasadas imperaba el paradigma funcionalista que resaltó el concepto de control experimental y que tuvo su mayor influencia en el desarrollo de los controles, la estandarización y las normas de comparación lo cual fue característico de la Teoría Clásica de los Tests - TCT (Lord y Novick, 1968; Cliff, 1979)

En la actualidad impera el paradigma Cognitivo signado por la priorización del concepto de proceso y su influencia en el campo de la medición se ha dado con el proceso de calibración de los ítems y el desarrollo de los llamados bancos de ítems, lo cual ha caracterizado a la Teoría de la Respuesta a los Item (TRI) o Teoría de la Estructura Latente (Jones y Appelbaum, 1989).

Esta teoría parte del supuesto que las puntuaciones obtenidas en un Ítem - y por ende en un test - por un sujeto dependen directamente del grado o nivel en que el sujeto posee la habilidad o rasgo medido. Este rasgo no es observado por lo que se denomina rasgo latente. La relación que existe entre el rendimiento del sujeto (puntuaciones observables) y el rasgo latente (habilidad del sujeto) puede ser descrita a través de una función matemática. Este rasgo latente puede corresponder a un constructo psicológico que permite explicar los hechos evaluados (Jones y Appelbaum, 1989).

Esta teoría, se origina cuando los investigadores en el campo de la psicometría se dan cuenta de las graves limitaciones que presenta la TCT siendo las principales: **a)** los resultados de los Instrumentos que miden diferentes constructos no pueden ser directamente comparables entre sí y, **b)** los instrumentos dependen de las muestras con la que se han trabajado, por lo tanto son relativos en el tiempo y las muestras de los ítems que conforman el test. Las características básicas de este afronte son **a)** invarianza del parámetro de la habilidad, implica que la estimación de la habilidad de un sujeto es independiente de la muestra de ítems que conforman el test. **b)** invarianza de los parámetros del ítem, implica que las características psicométrica de un tests como son la dificultad y la discriminación son independientes de la muestra específica utilizada para estudiar las características del ítem, y **c)** es posible obtener un estadístico que indica la precisión con la que se estima la habilidad o rasgo de cada sujeto por lo que puede variar de un sujeto a otro.

Este enfoque en la actualidad se ha desarrollado en dos vertientes: **a)** el enfoque Paramétrico que incluye al modelo de Rasch y los Modelos de 2, 3 y 4 parámetros, y es muy restrictivo pues exige que las puntuaciones correspondan a una formulación matemática que defina con precisión las CCI y **b)** el enfoque No paramétrico, que es menos restrictivo y por ende corresponde a las mediciones del tipo ordinal, entre los que destaca el modelo de Mokken.

En el campo de la construcción y evaluación de las pruebas psicológicas especialmente las de tipo genético, aún es limitado la aplicación de este enfoque enfoque habiéndose mayoritariamente trabajado con la metodología clásica que en este caso corresponde básicamente al modelo de escalamiento de Guttman.

El modelo de Guttman

El modelo de escalamiento de Guttman, es también conocido como el modelo de las escalas acumulativas o análisis del escalograma (Nunnally y Bemstein, 1995); Anastasi y Urbina, 1998). El propósito de la escala de Guttman es desarrollar un continuo unidimensional para un concepto que se desea medir. Formalmente la escala trata de predecir las respuestas de los sujetos a los ítems a partir del puntaje total obtenido. El objetivo del escalamiento es obtener un conjunto de ítems que conformen perfectamente este modelo, para verificar esto es necesario realizar el análisis del escalograma, que es un procedimiento diseñado para ordenar estímulos y sujetos respecto de alguna dimensión, por lo que constituye un modelo determinista de escalamiento, en el cual cada valor evaluado es una estimación de un continuo subyacente.

Esta técnica fue desarrollada por Guttman a partir de los trabajos hechos con los soldados americanos después de la II Guerra Mundial (Nunnally y Bernstein, 1995). Durante muchos años ha tenido una enorme difusión y ha sido considerada por muchos investigadores como el método básico para el escalamiento de opiniones y actitudes (Dotson y Summers, en Summers, 1986).

La idea básica de una escala de Guttman es que los sujetos y los estímulos pueden ser ordenados en un continuo, de tal forma que si un sujeto responde de forma correcta (o presenta una actitud positiva hacia un estímulo), dará respuestas correctas, con respecto a todos los estímulos que estén ordenados por debajo de éste; al contrario, si falla, o da respuesta negativa, a un estímulo determinado la dará también a todos los que estén ordenados por encima, tal como se puede apreciar en la **Tabla No.1** (Dawes, 1975).

Tabla No. 1
Matriz de una escala perfecta de Guttman

| | A | B | C | D |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 |
| 3 | 1 | 1 | 0 | 0 |
| 4 | 1 | 1 | 1 | 0 |
| 5 | 1 | 1 | 1 | 1 |

Los supuestos básicos del Modelo de Guttman (García Cueto, 1993) son:

- a. Definida una variable los sujetos deben ser ordenados con respecto a esta variable en un continuo unidimensional.
- b. Los estímulos utilizados pueden a su vez ser ordenados en un continuo de las mismas dimensiones
- c. Una respuesta positiva de un sujeto a un Ítem indica que ambos, Ítem y sujeto, están en el mismo punto del continuo.

El axioma básico del modelo, indica: que:

Si c_i y i son los puntos asociados al individuo i y al estímulo j , respectivamente entonces:

$$c_i \geq c_j \quad \text{si y solo si} \quad i \geq j$$

Según el modelo, si un sujeto supera una prueba j y falla otra prueba k , entonces:

$$q_k \geq c_i \geq q_j$$

De acuerdo a ésta lógica el individuo con una actitud más negativa o que peor discrimina un estímulo, da resultados negativos en todas las pruebas de la escala, el siguiente sujeto sólo superará la prueba más fácil, que signifique una actitud favorable pero muy débil, y así sucesivamente.

Según el modelo, a partir de la puntuación de cada sujeto en la escala se puede predecir correctamente la respuesta del sujeto a cada uno de los Ítems dicotómicos que componen la escala. Una puntuación de 2 indica que la persona en cuestión contestó positivamente a los dos Ítems más fáciles, o que implican una actitud menos favorable o un estado de ánimo más bajo, etc., de la escala y fracasó en todo el resto, o lo que es lo mismo, contestó de forma negativa al resto de Ítems de la escala.

Como indica la Bibliografía (Summers, 1986; Dawes, 1975; García Cueto, 1993 y Nunnaly y Bernstein, 1995), este patrón de respuesta resulta prácticamente imposible de encontrar en una aplicación real de cualquier escala. Este modelo es determinista y por lo general resulta imperfecto en la realidad. El principal problema consiste en establecer el grado de desviación con respecto al modelo ideal que se puede admitir para que el modelo sea útil como una adecuada representación de los datos empíricos, de manera que podamos decir que existe un ajuste adecuado de los datos al modelo.

El error es definido como la desviación de la respuesta observada del patrón ideal requerido por el modelo del escalograma de Guttman. Se asume que el monto de la desviación o del error observado es una función de la no adecuación del ítem o de los sujetos a los procedimientos utilizados.

Para medir el error, se utiliza la reproductibilidad, en una escala perfecta las respuestas que un sujeto da a todos los ítems pueden ser reproducidas exactamente por el mero hecho de conocer la posición del sujeto en el continuo de la escala. Es decir, si en una escala de 4 ítems un sujeto está en la posición 2 su patrón de respuestas necesariamente ha sido:(+ + - -), significando el signo + respuestas positivas y el - respuestas negativas. Cuando el sujeto da un patrón de respuestas como (+ + - +) su patrón de respuestas no es reproducible, de ahí que sea un error y por ende hay una falta de reproductibilidad. Con 3 respuestas positivas su patrón de respuestas ideal debería ser (+ + + -); pero al hacer esta transformación se supone que en el patrón empírico hay dos errores con respecto al patrón teórico, el elemento tres era negativo y debía ser positivo y el elemento 4 era positivo y debería ser negativo.

Para evaluar la relevancia del error en el escalograma se calcula el índice de reproductibilidad, cuya definición matemática es:

$$IR = 1 - \frac{\text{NÚMERO TOTAL DE ERRORES}}{\text{Número de ítems} * \text{Número de sujetos}}$$

Una escala de Guttman puede considerarse adecuada si se obtiene los siguientes resultados (Torgerson, 1958, en García Cueto, 1993):

- a. El índice de reproductibilidad de cada uno de los ítems es, al menos, de 0,85.
- b. El índice de reproductibilidad de la escala es de 0,90 o mayor.
- c. El esquema resultante de los errores no puede tener ningún tipo de interpretación que no sea la de deberse al azar.
- d. El tipo de respuesta más elegido de cualquier ítem no debe de ser elegido por más del 80 % de los sujetos, ya que si no el índice de reproductibilidad puede verse aumentado artificialmente debido a ese motivo.

Aplicaciones del Modelo de Guttman en el campo de las pruebas operatorias

El análisis de Guttman se utilizó en un principio exclusivamente en psicología social, su aplicación en la psicología infantil se produjo con la estandarización de las pruebas genéticas llevada a cabo por Inhelder y Vinh-Bang (Longeot, 1978).

Matalón (en Longeot, 1978) indica que la psicología genética es el campo más apropiado para la aplicación del análisis jerárquico, estableciendo que cada estadio puede describirse por una estructura que integra todas las adquisiciones de los estadios precedentes. En este sentido, los diferentes estadios del desarrollo constituyen una escala. Los ítems del estadio formal, cuya estructura operatoria es genéticamente posterior a la estructura operatoria concreta, no pueden realizarse con éxito antes que los ítems del estadio concreto. Por el contrario, si los problemas del estadio concreto son resueltos con éxito, los problemas del estado formal, que es posterior, pueden ser resueltos positivamente o no, según que los niños hayan alcanzado ya el estadio formal o se encuentren todavía en el estadio concreto. Ocurre igual con los subestadios sucesivos de un mismo estadio. Motivo por el cual existe una estrecha relación con el modelo de Guttman.

El uso de este método, es posible en psicología genética, desde el momento en que se dispone de una teoría de las fases sucesivas de desarrollo, de modo que la teoría fija el orden según el cual se jerarquizan los ítems y los sujetos sólo pueden ser clasificados según su puntuación total. De esta manera los ensayos y la arbitrariedad se reducen al mínimo e incluso se suprimen por completo. Las técnicas más utilizadas han sido las de Goodenough y la de Green (Longeot, 1978).

El estudio de las respuestas correctas e incorrectas dadas por los sujetos a los ítems ordenados de esta forma permite verificar o invalidar una hipótesis genética. El estudio de la jerarquía de las respuestas permite juzgar si las pruebas diferenciales que se construyen están en correspondencia o no con la teoría que las sustenta.

El modelo de Mokken

El desarrollo de las escalas Mokken, se inicia en 1971, cuando este autor publica un libro en el que expone su teoría y procedimientos de construcción de escalas para ítems dicotómicos. Esta teoría es considerada la versión probabilística de las escalas Guttman para ítems dicotómicos, así como la aproximación más general no paramétrica de la teoría de la Respuesta al Ítem (TRI), aunque Molenaar (1982 y 1986) ha sugerido una ampliación del modelo de Mokken para ítems dicotómicos para el caso de los ítems politómicos.

Muchos constructos psicológicos son medidos por medio de Ítems debido a la imposibilidad de ser medidos directamente, en estos casos la ubicación del sujeto en el

Rasgo Latente (Hipotético Atributo) es inferido a través de las respuestas presentadas a los ítems.

La teoría de la respuesta al ítem (TRI) explica las respuestas observadas asumiendo que existe un continuo en el cual los ítems y los sujetos pueden ser ubicados. En este caso la probabilidad de emitir una respuesta está expresada en función del atributo que es medido y de ciertas características de los ítems en lo que respecta a su dificultad y poder de discriminación (Mokken & Lewis, 1982).

Este afronte no hace uso de ninguna función matemática específica, por ello es en consecuencia un modelo empírico que brinda información solamente de tipo ordinal obtenida a partir de la ubicación de los Ítems y los sujetos con relación al continuo estudiado (Mokken, Lewis & Sijtsma, 1982). Los principales supuestos teóricos: del Modelo (Mokken & Lewis, 1982) son:

El modelo asume que (n) individuos responden a (k) Ítems dicotómicos, en los cuales una respuesta es positiva y la otra es negativa con respecto al rasgo latente que se estudia. Los supuestos teóricos (Mokken & Lewis, 1982) son:

- a. *Dicotomía de los ítems*, la variable manifiesta que se mide, deber ser presentada de forma binaria, ya sea (V o F), acierto o error, aunque es posible generalizar el modelo a ítems politómicos.
- b. *Monotonicidad*, para cada ítem la probabilidad emisión de una respuesta positiva de los sujetos se incrementa de forma (Θ) estrictamente monótonica, la función que describe la relación entre esta probabilidad y el rasgo latente es denominada la Curva Característica del Ítem (CCI) (Debets, Brouwer, Sijtsma & Molenaar, 1989). Los sujetos pueden ser ubicados bajo el continuo latente y los ítems también pueden ser ubicados en el mismo continuo de acuerdo a su dificultad (δ) esto significa que para ciertos ítems la relación con el rasgo latente debe ser mayor y por lo tanto deben tener más respuestas positivas.

La dificultad de un ítem es definida como el valor alcanzado por un sujeto con respecto a una respuesta positiva con probabilidad igual a de 0.50; excepto para el caso de la fluctuaciones muestrales, en la muestra el orden de la dificultad (δ) corresponde al orden inverso de la proporción de sujetos que responde positivamente a cada Ítem. El orden de los sujetos esta basado en el número de respuestas positivas, si un grupo de ítems tiene una CCI monótonicamente creciente, cada medida del mismo atributo es denominado como un grupo de ítems homogéneos y monótonos, y ello se muestra cuando el orden esperado de los examinados en el rasgo latente es el mismo para cada selección de ítems que conforman un grupo homogéneo.

- c. *Unidimensionalidad*, todos los indicadores miden el mismo rasgo latente y que cada sujeto tiene un valor (Θ) desconocido.
- d. *Independencia local estocástica*, para las personas con un valor (Θ) en las respuestas de un grupo de ítems, estos deben ser estocásticamente independientes.
Las CCI no se deben intersectar, un ítem que tiene una alta dificultad con relación a otro ítem, tiene una baja probabilidad de presentar respuestas positivas. El grupo de ítems es llamado doblemente monótono cuando el ítem es libre del orden de los sujetos, lo que se demuestra cuando el orden de los ítems según el nivel de dificultad es el mismo para cada sub-población de personas de la población en la que el modelo de doble monotonicidad es construido. Es factible que en el modelo de Mokken un sujeto responda positivamente a un Ítem de alta dificultad y de una respuesta negativa a un Ítem de baja dificultad, ello se debe a que este es un modelo de tipo probabilístico.

Análisis estadístico de los ítems

- a. Escalabilidad de dos ítems (i,j) puede ser estimada o contrastada por medio del coeficiente de escalabilidad de Loevinger (h_{ij}), que es definido, como:

$$H_{ij} = 1 - f_{ij}/e_{ij}$$

Donde:

f_{ij} , es el número observado de errores de acuerdo al modelo de la escala de Guttman.

E_{ij} , es el número esperado de errores, asumiendo que las respuestas a los ítems sean independientes de las personas y que los marginales sean fijos.

- b. *Escalabilidad de un ítem* (h_i) con relación a los otros ítems del test y puede ser calculado por medio de la siguiente fórmula:

$$h_i = 1^k - \sum^k f_{ij} / \sum e_{ij}$$

- c. *Escalabilidad de un grupo de ítems*, se calcula a través de:

$$H = \sum \sum f_{ij} / \sum \sum e_{ij}$$

Para estos coeficientes de escalabilidad, Mokken ha desarrollado la Teoría del Muestreo Asintótico (Kingma & Ten Vergert, 1985; Kingma & Reuvekamp, 1986); introduciendo el estadístico Delta star (Δ^*) que tiene una aproximación a la distribución standard normal, este estadístico permite evaluar la hipótesis que en la población:

$$H: \Delta^*=0 \text{ (o su equivalente } h=0)$$

Este coeficiente puede ser utilizado para la evaluación de un grupo fijo de ítems o para construir una escala de un grupo de ítems.

En el análisis de las escalas Mokken, para un grupo de ítems se define:

- $h_{ij} > 0$ para todos los pares de ítems.
- $H_i > =c$ (además $h \geq c$) para cada ítem

En esta definición c es una constante que puede ser seleccionada por el constructor de pruebas, pues Mokken recomienda elegir entre 0.30, 0.40 ó 0.50 (Debets & Brouwer, 1989). Adicionalmente es posible además formular otros 2 criterios para evaluar la adecuación de la escala, los cuales son:

- a. Los coeficientes h y h_i , deben ser estadísticamente mayores que 0. Esta propiedad requiere que todos los h_{ij} sean positivos.
- b. El trazado de las líneas debe ser doblemente monótono, aunque esta propiedad no es básica para la definición de la escala pues es una restricción adicional para los datos.

Mokken basándose en los coeficientes de escalabilidad ha desarrollado 2 procedimientos diferentes para construir escalas (Kingma & Taerum, 1989), los cuales son: a) Evaluar los ítems de una escala a-priori, y b) Construir de manera escalonada (stepwise) una o más escalas de un grupo grande de ítems.

Tal como indican Kingma & Taerum, 1989 la construcción de una escala mediante el procedimiento de búsqueda se basa únicamente en una definición matemática de la escala y por ello no es posible poder sustituir el análisis del significado teórico de los valores del instrumento en la investigación psicológica.

Los coeficientes de escalabilidad y la definición de la escala están relacionados con un grupo de ítems monótonos y homogéneos. Otra evaluación heurística del modelo de Monótona Homogeneidad es propuesto por Molenaar (1986) quien indica que para cada ítem la probabilidad de una respuesta positiva debe ser monotonamente no creciente con el incremento de los puntajes, como esta revisión no puede hacerse sobre bases matemáticas, debe ser aplicada con precaución.

Cuando se asume adicionalmente la doble monotonicidad del grupo de ítems, ésta puede ser revisada por medio de la llamada matriz p en la que cada ítem es ordenado de acuerdo a su nivel de dificultad.

En esta matriz para cada par de ítems la probabilidad de emitir una respuesta positiva en todos los ítems es presentada, si en el trazado de las líneas no se intersectan, en la matriz p las probabilidades deben incrementarse de izquierda a derecha y de arriba hacia abajo.

Estimación de la confiabilidad

La estimación de la confiabilidad de los puntajes verdaderos de las escalas Mokken, esta basada en el modelo de doble motonocidad, la cual es similar a la presentada por la teoría Clásica de los Test (TCT) (Debets & Brouwer, 1989) y que puede ser formulada como:

$$\sigma^2_{xx'} = \sum^k \sum^k (\pi_{ij} - \pi_i \pi_j) / \sigma^2(x)$$

La varianza de los puntajes del test σ^2 puede ser estimada directamente de los datos.

La dificultad de la población π_i ($i = 1, \dots, k$) Puede ser estimada de la proporción de sujetos en la muestra que responden positivamente los ítems. Los pares de dificultad de la población π_{ij} para los diferentes ítems pueden ser estimados con la proporción de respuestas positivas en los ítems. Es necesario estimar los elementos π_{ij} : como la proporción de los sujetos que tienen una respuesta positiva en el mismo ítem en dos pruebas independientes, esta proporción no puede ser estimada en base a la proporción de la muestra.

Lamentablemente el modelo de Mokken a pesar de sus interesantes propiedades como método de construcción de pruebas psicológicas ha tenido escasa difusión y nula publicaciones en castellano tal como lo afirman Rivas y Martínez (1992).

Aplicaciones del Modelo de Mokken en el campo de las pruebas operatorias

Una de las principales aplicaciones del modelo de mokken en el campo de las prueba operatorias piagetanas, es la realizada por Kingma & Reuvekamp (1984) quienes construyeron una escala para evaluar el desarrollo de la seriación en niños, la cual en un estudio posterior fue validada por Kingma & Loth (1985). También se observa que De Jong y Molenaar (1987) desarrollaron un estudio en el cual utilizaron el modelo de Mokken para construir una escala de Clasificación psiquiátrica de las disfunciones

sociales.

Objetivos del estudio

1. Evaluar las características psicométricas TO.F.C. de Longeot según el Modelo de Guttman.
2. Evaluar las características psicométricas del test de operaciones formales combinatorias (TO.F.C.) de Longeot según el modelo de Mokken.
3. Comparar las características psicométricas del TO.F.C. de Longeot obtenidas con el modelo de Guttman en relación a las obtenidas con el modelo de Mokken.

MATERIAL Y METODOS

1. Diseño de Investigación.

El diseño de la investigación corresponde al Descriptivo correlacional en la medida que los resultados obtenidos en la medición del fenómeno son utilizados para analizar las características psicométricas del instrumento utilizado haciendo uso de los coeficientes de correlación (Alarcón, 1991, Sánchez y Reyes, 1998; Hernández, Fernández y Baptista, 1998).

2. Participantes de la Investigación

El universo de investigación estuvo conformado por los alumnos regulares Matriculados en la Facultad de Psicología de la U.N.M.S.M., que alcanzó en 1998 a 686 casos. En la presente investigación se trabajó con el 30% de la población total que corresponde a 260 casos distribuidos de forma representativa en las promociones entre 1994 y 1998.

Para la obtención de la muestra se aplicó un procedimiento de muestreo no probabilístico por cuotas, clasificando a los alumnos de acuerdo a las promociones y los turnos en que se encontraban matriculados.

3. Instrumento de medición.

El test de operaciones formales combinatorias (T.O.F.C.), forma parte de los test Operatorios Colectivos de Longeot y comprende 7 ítems correspondientes a los tres tipos de operaciones combinatorias que Inhelder y Piaget (1951, en Longeot, 1978) han investigado en su estudio de la génesis de la idea de azar en el niño: las combinaciones, en las cuales deben descubrirse todas las parejas posibles con un número dado de elementos, las permutaciones, en las que deben descubrirse todos los órdenes posibles de presentación de un número dado de elementos y los ordenamientos, síntesis de las dos operaciones anteriores, en las que deben descubrirse a un tiempo las parejas y las

permutaciones. Los niños escriben sus respuestas en la hoja que para cada ítem tiene siempre un número de líneas superior al que necesita. El test consta de tres niveles:

- a. El primero está constituido por los dos problemas de ordenamientos y un problema de simple multiplicación lógica de dos subconjuntos (las parejas de un baile, formadas entre bailadores y bailadoras).
- b. El segundo por el problema de las combinaciones (ping-pong) y las dos preguntas en las que se pide calcular el número de combinaciones y de ordenamientos que se obtienen al añadir un elemento, lo que supone el descubrimiento de la ley. Estas dos preguntas cuentan como un solo ítem.
- c. El tercero reúne los dos problemas de permutaciones con cuatro elementos, ya que difícilmente pueden ser resueltos sin la utilización del método riguroso consistente en bloquear, al principio de cada línea, los dos mismos elementos y cambiar la posición de los otros dos.

El T.O.F.C., tiene la ventaja de permitir un examen clínico de las respuestas. El método empleado por el sujeto para escribir combinaciones sobre la hoja es, en efecto, característico del estadio

Investigaciones antecedentes del test de Longeot

En nuestro medio el TOFC fue aplicado por Majluf (1974) en su estudio sobre el desarrollo del pensamiento formal proposicional y combinatorio de dos grupos de adolescentes de diferentes estratos socio-económicos de Lima. En tanto que Bello y Mejía (1993) estudiaron las relaciones entre el nivel de desarrollo moral e intelectual, edad y género en alumnos de secundaria de un colegio experimental alternativo de Lima.

Entre las investigaciones relacionadas en el extranjero, encontramos que Lis & Magro (1993), aplicaron el test de pensamiento formal operatorio de Longeot en un grupo de adolescentes italianos.

Para el desarrollo de presente investigación el T.O.F.C. fue adaptado utilizando la siguiente estrategia:

- a. Adaptación Lingüística de los ítems, revisando el sentido de las preguntas formuladas de manera que se realizaron las correcciones para hacerlas más comprensible a los participantes.
- b. Análisis de los ítems de acuerdo al Modelo del Escalograma de Guttman, revisando los patrones de respuesta y el coeficiente de reproducibilidad.
- c. Análisis de los Ítems de acuerdo al Modelo de Mokken, calculándose la dificultad de cada ítem, los coeficientes de escalabilidad de Loevinger y el coeficiente delta star para evaluar su significatividad.

- d. Evaluación comparativa del ajuste de los ítems a los dos modelos.
- e. Estudio de la validez, en base al estudio de la validez de Construcción, utilizando el método intraprueba y el análisis factorial exploratorio.
- f. Estudio de la confiabilidad, en base a la teoría clásica de los tests, a través del coeficiente Kuder Richardson 20.

RESULTADOS

Los resultados encontrados permiten observar que en el análisis de los ítems del Test T.O.F.C. en base al modelo de Guttman se presentan patrones de respuestas difusos, tal como se aprecia en la tabla 2, lo cual puede haber sido la causa para que su coeficiente de reproductibilidad asciende solamente a 0.53, lo que equivale a obtener un instrumento que no cumple con los niveles de exigencia requeridos para conformar una escala Guttman que sea consistente. Este hallazgo nos permite concluir que el T.O.F.C. no se adecua a una escala determinista de tipo Guttman.

En relación al análisis de la dificultad de los ítems presentado en la *Tabla No. 3*, podemos notar que exhiben un patrón de incremento progresivo de la dificultad, lo cual resulta indispensable en toda prueba operatoria.

En relación al análisis realizado en base al Modelo de Mokken, presentado en la *Tabla No. 4*, se puede apreciar que todos los coeficientes $H(i,j)$ de Loevinger entre los ítems, son positivos y fluctúan entre 1.00 y 0.22, lo cual es aceptable para el modelo estudiado.

Tabla 3
Niveles de Dificultad de los Ítems del T.O.F.C.

| Variable | Dificultad |
|----------|------------|
| Item 1 | 0.87 |
| Item 2 | 0.85 |
| Item 3 | 0.82 |
| Item 4 | 0.65 |
| Item 5 | 0.58 |
| Item 6 | 0.51 |
| Item 7 | 0.29 |
| Item 8 | 0.17 |

Tabla 2
Análisis de los ítems del T.O.F.C. a través del Modelo de Guttman

| PATRON DE RESPUESTA | Casos | % |
|---|---------------|---------------|
| 0 0 0 0 0 0 0 0 | 8 | 3.1 |
| 1 0 0 0 0 0 0 0 | 8 | 3.1 |
| 1 0 1 0 0 0 0 0 | 6 | 2.3 |
| 1 0 0 1 0 0 0 0 | 3 | 1.2 |
| 1 0 0 0 1 0 0 0 | 2 | .8 |
| 1 1 0 0 0 0 0 0 | 8 | 3.1 |
| 1 1 0 0 0 1 0 0 | 1 | .4 |
| 1 0 0 1 0 1 0 0 | 1 | .4 |
| 1 0 1 1 0 0 0 0 | 5 | 1.9 |
| 0 1 0 1 1 0 0 0 | 1 | .4 |
| 0 1 1 1 0 0 0 0 | 6 | 2.3 |
| 0 0 1 1 0 1 0 0 | 4 | 1.5 |
| 1 1 0 0 1 0 0 0 | 1 | .4 |
| 1 1 0 0 0 1 0 0 | 1 | .4 |
| 1 0 1 0 1 0 0 0 | 2 | .8 |
| 1 0 1 1 1 0 0 0 | 3 | 1.2 |
| 1 0 1 1 0 1 0 0 | 2 | .8 |
| 1 0 0 1 1 1 0 0 | 1 | .4 |
| 0 1 1 1 1 0 0 0 | 1 | .4 |
| 0 1 1 1 1 1 0 0 | 1 | .4 |
| 1 1 0 0 1 1 0 0 | 1 | .4 |
| 1 0 1 1 1 1 0 0 | 3 | 1.2 |
| 1 1 0 1 0 0 0 0 | 3 | 1.2 |
| 1 1 0 1 0 1 0 0 | 1 | .4 |
| 1 1 0 1 1 0 0 0 | 1 | .4 |
| 1 1 0 1 1 1 1 1 | 8 | 3.1 |
| 1 1 1 0 0 0 0 0 | 39 | 15.0 |
| 1 1 1 0 0 1 1 1 | 1 | .4 |
| 1 1 1 0 1 0 0 0 | 11 | 4.2 |
| 1 1 1 0 1 1 0 0 | 8 | 3.1 |
| 1 1 1 0 1 1 0 1 | 2 | .8 |
| 1 1 1 1 0 0 0 0 | 13 | 5.0 |
| 1 1 1 1 0 1 0 0 | 4 | 1.5 |
| 1 1 1 1 0 1 1 1 | 2 | .8 |
| 1 1 1 1 1 0 0 0 | 11 | 4.2 |
| 1 1 1 1 1 0 0 1 | 1 | .4 |
| 1 1 1 1 1 1 0 0 | 23 | 8.8 |
| 1 1 1 1 1 0 0 1 | 30 | 11.5 |
| 1 1 1 1 1 1 1 0 | 36 | 1.2 |
| 1 1 1 1 1 0 1 1 | 4 | 1.5 |
| 1 1 1 1 1 1 1 1 | 26 | 10.0 |
| TOTAL | 260 | 100.00 |
| Coefficiente de Reproducibilidad | = 0.53 | |

Tabla 4
Matriz de coeficiente H(i,j) de Loevinger entre ítems del F.O.F.C. de Longeot

| VARIABLES | ITEM8 | ITEM7 | ITEM6 | ITEM5 | ITEM4 | ITEM3 | ITEM2 |
|-----------|-------|-------|-------|-------|-------|-------|-------|
| ITEM7 | 0.90 | | | | | | |
| ITEM6 | 0.81 | 0.86 | | | | | |
| ITEM5 | 0.84 | 0.87 | 0.60 | | | | |
| ITEM4 | 0.97 | 0.92 | 0.62 | 0.43 | | | |
| ITEM3 | 1.00 | 0.96 | 0.47 | 0.56 | 0.30 | | |
| ITEM2 | 0.03 | 0.37 | 0.36 | 0.40 | 0.41 | 0.30 | |
| ITEM1 | 1.00 | 1.00 | .50 | 0.74 | 0.33 | 0.42 | 0.22 |

N = 260

Respecto al análisis de los coeficientes Delta Star (i,j) de los ítems, (ver *Tabla No.5*), que permiten establecer la significación estadística de los ítems en el modelo,

Tabla 5
Matriz de coeficiente Delta Star(i,j) entre los ítems del F.O.F.C. de Longeot

| VARIABLES | ITEM8 | ITEM7 | ITEM6 | ITEM5 | ITEM4 | ITEM3 | ITEM2 |
|-----------|-------|-------|-------|-------|-------|-------|-------|
| ITEM8 | 14.63 | | | | | | |
| ITEM6 | 8.24 | 11.99 | | | | | |
| ITEM5 | 7.29 | 10.43 | 11.63 | | | | |
| ITEM4 | 7.28 | 9.55 | 10.22 | 8.26 | | | |
| ITEM2 | 4.80 | 6.37 | 5.01 | 6.88 | 4.27 | | |
| ITEM3 | 2.16 | 2.44 | 3.75 | 4.88 | 5.84 | 6.68 | |
| ITEMI | 4.03 | 5.56 | 4.43 | 7.67 | 4.00 | 7.83 | 4.14 |

N= 260

Por otro lado en la *Tabla No. 6* se presentan los resultados de los coeficientes H de Loevinger, y se aprecia que en todos los casos son estadísticamente significativos ($p < .05$), cosa que igualmente ocurre en el caso de la escala total que asciende a 0.81. estos hallazgos permiten concluir que el T.O.F.C. de Longeot corresponde a una prueba que cumple con los supuestos del Modelo de Mokken. También se observa que la estimación de la confiabilidad no es alta, ascendiendo a 0.82

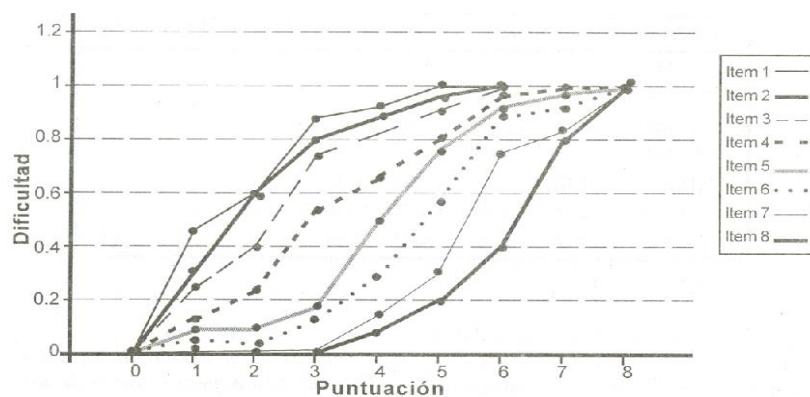
Tabla 6
Análisis de la Escalabilidad del T.O.F.C. de Longeot a través de los Coeficientes H de Loevinger

| Variable | H(i) | Delta (i) |
|--------------|----------|-----------|
| Item 1 | 0.49 | 14.15 * |
| Item 2 | 0.32 | 10.47 * |
| Item 3 | 0.47 | 15.48 * |
| Item 4 | 0.53 | 19.21 * |
| Item 5 | 0.60 | 21.91* |
| Item 6 | 0.62 | 21.68 * |
| Item 7 | 0.83 | 23.49 * |
| Item 8 | 0.86 | 18.14 * |
| Escala Total | H = 0.81 | 36.65 |

Confiabilidad no = 0.82
N= 260 * $p < .05$

Adicionalmente el análisis gráfico de las curvas características de los ítems presentados en el *Gráfico No. 1*, permite observar que las curvas características de los

ítems son paralelas y por lo tanto no se intersectan, lo que permite verificar el cumplimiento del supuesto de doble monotonicidad que debe existir entre los ítems de instrumento que corresponde al modelo de Mokken.



Respecto al análisis psicométrico complementario de la validez de construcción del instrumento efectuado en base al método intra-prueba, utilizando el análisis factorial exploratorio en base al procedimiento de los componentes principales y la rotación Varimax, en la *Tabla No. 7*, podemos observar que los resultados indican que existe un sólo factor en el conjunto de ítems evaluados. Este factor permite explicar el 68.43% de la varianza total, lo que nos permite concluir que el *T.O.F.C.* presenta validez de construcción.

Tabla 7
Análisis Factorial Exploratorio del T.O.F.C. de Longeot

| Variable | Factor 1 |
|----------|----------|
| Item 1 | .83 |
| Item 2 | .85 |
| Item 3 | .79 |
| Item 4 | .69 |
| Item 5 | .76 |
| Item 6 | .80 |
| Item 7 | .94 |
| Item 8 | .93 |

Varianza Explicada 68.43%
N = 260

La revisión del análisis de ítems a través del coeficiente de correlación ítemtest indica que se alcanzan correlaciones superiores a 0.28 las cuales son estadísticamente significativas ($p < .05$), encontrándose además que la confiabilidad por consistencia interna en base al coeficiente Kuder-Richardson 20, alcanza un coeficiente de 0.74, lo que determina que el test es confiable bajo los supuestos de la Teoría Clásica de los Tests.(Ver *Tabla No. 8*)

Tabla 8
Análisis de Ítems y Con fiabilidad del T.O.F.C.

| Variable | ritc |
|----------|-------|
| Item 1 | .28 * |
| Item 2 | .35 * |
| Item 3 | .41 * |
| Item 4 | .52 * |
| Item 5 | .64 * |
| Item 6 | .29 * |
| Item 7 | .34 * |
| Item 8 | .57 * |

Kuder Richardson 20 = 0.74 *

N = 260 * significativo $p < .05$

Con respecto a los resultados generales obtenidos en los participantes, en la *Tabla No. 9* se puede apreciar que 100 casos (38.50 %) se ubica en el nivel de las operaciones formales combinatorias Concretas, 87 casos (33.50 %) en el nivel de las operaciones formales combinatorias tipo A y 73 casos (28.00 %) en el nivel de las operaciones formales combinatorias tipo B.

Tabla 9
Niveles del desarrollo alcanzado en el T.O.F.C.

| NNEL | Casos | % |
|--------------|------------|---------------|
| Concreto | 100 | 38.50 |
| Formal A | 87 | 33.50 |
| Formal B | 73 | 28.00 |
| TOTAL | 260 | 100.00 |

El análisis comparativo según el sexo de los alumnos, presentado en la *Tabla No. 10*, permite observar que no existen diferencias estadísticas significativas ($U = 8043.50$, $P > .05$), es decir que ambos grupos presentan niveles similares de funcionamiento en lo que corresponde a las operaciones formales combinatorias .

Tabla 10
Análisis Comparativo del T.O.F.C. según el sexo a través de la U de Mann- Whitney

| T.O.EC | Sexo | | U | Z |
|--------------|----------------------------|----------------------------|----------------|--------------|
| | Varón N= 130 M rango | Mujer N= 130 M rango | | |
| Ptje. | 132.54 | 128.98 | 8043.50 | -0.40 |

* $p < .05$

Finalmente en la comparación por promoción a la que pertenecen los alumnos, (Ver *Tabla No. 11*), encontramos que el análisis de Varianza ordinal de Kruskal-Wallis indica que no existen diferencias estadísticas significativas ($X^2 = 3.13$, g.l. = 4, $p = .2092$). Es decir que en todas las promociones se presentan niveles similares de funcionamiento en las operaciones formales combinatorias.

Tabla 11
Análisis Comparativo del T.O.F.C según Promoción a través del Análisis de Varianza Ordinal de Kruskal-Wallis

| Promoción | | | | | Chi-Cuadrado | P |
|-------------------|-------------------|-------------------|-------------------|-------------------|--------------|--------------|
| 94 | 95 | 96 | 97 | 98 | | |
| N = 52 M rango | N = 52 M rango | N = 52 M rango | N = 52 M rango | N = 52 M rango | | |
| 126.30 | 130.18 | 135.12 | 140.20 | 148.23 | 3.13 | .2092 |

* $p < .05$

DISCUSION

Los resultados encontrados han permitido demostrar que el test T.O.F.C. presenta un mejor ajuste al modelo de Mokken que al modelo de Guttman, lo cual concuerda con los estudios de Kingma & Reuvekamp (1984) Y Kingma & Loth (1985) en trabajos realizados con una prueba de seriación.

Estos hallazgos implican que para las mediciones psicológicas de variables operatorias o genéricas, es más relevante asumir un enfoque del tipo probabilístico que un enfoque del tipo determinístico, debido a que cuando se realiza la medición de un constructo psicológico es necesario tomar en cuenta un margen de error probable que lo asemeja más a la realidad que el considerar una estructura libre de error, que puede ser poco probable de ocurrir en la realidad común.

La demostración de la efectividad del modelo de Mokken sobre el modelo de Guttman permite recomendar que en aquellos casos en los cuales es necesario contar con un instrumento psicológico que genere puntuaciones ordinales y que sea unidimensional, es más aconsejable utilizar esta metodología por ser más actual y eficiente.

Por otro lado el test T.O.F.C. al presentar, validez, confiabilidad y adecuación al modelo de Mokken, y por ende cumplir con los requisitos psicométricos básicos, puede constituirse en un instrumento valioso que puede ser utilizado para desarrollar nuevas investigaciones cognitivas en nuestro medio.

Los resultados obtenidos en los niveles de desarrollo cognitivo de las operaciones Formales combinatorias alcanzados por los participantes, nos indican que en la actualidad coexisten todos los niveles de desarrollo, es decir Niveles Concretos, Niveles Formales tipo A y los Niveles Formales tipo B, lo cual podría traducirse en cierta dificultad para trabajar con problemas que impliquen el uso de ese tipo de pensamiento y que se puede traducir en un rendimiento diferencial en áreas como las estadísticas y el análisis lógico entre otros.

BIBLIOGRAFIA

Alarcón, R. (1991). *Métodos y diseños de Investigación del Comportamiento*. Lima: U.P.C.H., Fondo Editorial.

Anastasi, A. & Urbina, S. (1998). *Test Psicológicos*. México: Prentice-Hall.

Bello, M. y Mejía, E. (1993). Relaciones entre el nivel de desarrollo moral e intelectual, edad y género en alumnos de secundaria de un colegio experimental / alternativo de Lima. *Psicología*. Vol. XI, N°.1, pp. 13-30.

Cliff, N. (1979). *La Psicometria, en Wolman B. Manual de Psicología General: Historia, teoría y método*, Barcelona: Ed. Martinez Roca S.A., Vol 1.

Dawes R. (1975). *Fundamentos y Técnicas de Medición de Actitudes*. México: Ed. Limusa.

De Jong, A. & Molenaar, I. (1987). An application of Mokken's model for stochastic, cumulative scaling in psychiatric research. *Journal of psychiatric Research*. Vol, 21, No. 2, pp.137-149

Debets, P. & Brouwer, E. (1989). *MSP: Mokken Scale Analysis for Polychotomus Items*, Groningen - The Netherlands: IEC-Programma,

Debets, P. & Brouwer, E., Sijstma, K. & Molenaar, I. W. (1989) MSP: A computer program for item analysis according to a Nonparametric IRT approach, *Educational And Psychological Measurement*, Vol. 49, Pgs. 609-613.

García Cueto (1993). *Introducción a la psicometría*. Madrid: Siglo XXI Editores.

Hernandez, R., Hernández, C., Baptista, L. (1998). *Metodología de la investigación*. México: Mc. Graw-Hill.

Jones, L. & Appelbaum, M. (1989). Psychometric Methods. *Annual Review of Psychology*. Vol. 40. Pgs. 23-43

Kingma, L. & Loth, F. L. (1985) The Validation of a developmental Scale for Seriation, *Educational And Psychological Measurement*, Vol. 44, Pgs. 321 - 328.

Kingma, J.. & Reuvekamp, J. (1984) . The construction of a developmental Scale for Seriation, *Educational And Psychological Measurement*, Vol. 44, Pgs. 1 - 21.

Kingma, J. & Reuvekamp, J. (1986) Mokken Scale: A Pascal program for Nonparametric Stochastic Scaling, , *Educational And Psychological Measurement*, Vol. 46, Pgs. 667 - 677.

Kingma, J. & Reuvekamp, J. (1986) Mokken Test for the robustness Nonparametric Stochastic Mokken Scales, , *Educational And Psychological Measurement*, Vol. 46, Pgs. 679 -685.

Kingma, J. & Reuvekamp, T. (1986) Spss- X Procedure and standalone Programs For The Mokken Scale Analysis: A nonparametric Item Response Theory Model, , *Educational And Psychological Measurement*, Vol.49,Pgs.101-136.

Kingma, J. & Ten Vergert E. M. (1985) A Nonparametric Scale Analysis of the Developmental of Conservation, *Applied Educational Measurement*, Vol. 9, No. 4, Pgs. 375 -387.

Lis, A. & Magro T. (1993). Study of Longeot's Test of formal operational thinking in a group of Italian adolescents. *Perceptual and Motor Skills*. Vol. 76, pp. 739-752.

Longeot, F. (1978). *Psicología diferencial y teoría operatoria de la inteligencia*. Barcelona: Edit. Omega.

Lord, F. & Novick, M. (1968). *Statistical Theories of mental tests scorer*. Massachusetts: Addison Wesley.

Majluf, A, (1974). *Desarrollo del pensamiento formal proposicional y combinatorio de dos grupos de adolescentes de diferentes estratos socio-económicos de Lima, Perú*. Tesis de Bachiller en Psicología. UNMSM.

Mokken, R. J. & Lewis, CH. (1982) A Nonparametric approach Analysis of Dichotomous item response, *Applied Psychological Measurement*, Vol 6, No. 4, Pgs 417-430.

Mokken, R. J., Lewis, CH. & Sijstma, K. (1986) Rejoinder to The Mokken Scale: A Critical Discussion, *Applied Psychological Measurement*, Vol 10, No. 3, Pgs 279 - 285 .

- Molennar, I. W. (1982), Some improved diagnostics for failure of the rasch model. *Psychometrica*, Vol 48, N° 1, Pp. 49-71.
- Molennar, I.W. (1986). *PML: users manual, pc version*. Groningen, The Netherlands: IEC proGAMMA
- Nunnally, J. & Bemstein, I. H. (1995). *Teoría psicométrica*. México: McGraw-Hill.
- Rivas, M.y Martínez T. (1992). MOKPAS: un programa para el escalamiento de ítems según el modelo de TRI no paramétrico Mokken. *Revista de Investigaciones psicológicas*. Num. 11. pp. 187-205.
- Sánchez, C. H. y Reyes, M. C. (1990). *Metodología y Diseños en la Investigación Científica: Aplicadas a la Psicología, Educación y Ciencias Sociales*. Lima: INIDE.
- Summers D. (1986). *Medición de Actitudes*. México: Ed. Trillas.