

ESTUDIO DE LAS CARACTERISTICAS PSICOMETRICAS
DE LA FORMA CORTA DEL TEST DE BEERY
CON EL MODELO RASCH

Luis Miguel Escurra Mayaute¹
Ana Esther Delgado Vásquez²
Andrés Aparcana Carrasco³

Se realizó el estudio de las principales características de la forma corta del Test de Desarrollo Visomotor de Beery, desarrollada por Salazar y Delgado en 1985, bajo los supuestos del modelo de Rasch.

Se aplicó la prueba a 1210 sujetos de Lima Metropolitana, (49.9% mujeres y 50.1% hombres) pertenecientes a centros educativos estatales y no-estatales, de 1º, 2º y 3º grado de primaria y que residen en zonas urbanas y urbanas-marginales.

Los resultados confirman que la prueba conformada por 12 ítemes al ser analizadas bajo el modelo probabilístico de Rasch obtienen valores que no se adecuan al modelo, en tanto que al eliminar 6 ítemes se conforma una escala que cumple con todos los supuestos del modelo. Se computaron los valores de los ítemes-parámetros y los persona-parámetros, así como los análisis estadísticos básicos para evaluar la bondad de ajuste al modelo.

Se estudió la validez del modelo al compararlo según las diferentes características demográficas estudiadas, constatándose la generalidad y estabilidad de la escala obtenida.

Adicionalmente se evaluó la relación existente entre la escala Rasch y la prueba obtenida originalmente.

Using the Rasch Model a study about the principle characteristics of the Short Form of Beery Visual-Motor Integration Test developed by Salazar and Delgado (1985) was carried out.

The test was administered to 1210 students (49,9% girls and 50,1% boys) coming from public and private schools of Lima from 1st, 2nd and 3rd grades.

The results show that the Short Form of the VMI constituted by 12 items, did not fit the Rasch model. However, when six items were eliminated a scale that fitted the model was constituted. Scores of the item-parameters and the person-parameters, as well as basic statistic analysis to evaluate the adjustment of the model, were computed.

The validity of the model was compared according to the different demographic characteristics, verifying the generability and stability of the scale. Additionally the relation between the Rasch Scales and the Short Form was evaluated.

-
1. Psicólogo, Profesor Auxiliar de la Facultad de Psicología, y Jefe de la Unidad de Cómputo del Instituto de Investigaciones Psicológicas de la Universidad Nacional Mayor de San Marcos y Docente de la Pontificia Universidad Católica del Perú. Realiza investigaciones en el campo de la Psicometría y Análisis de Datos.
 2. Psicólogo, Docente de la Pontificia Universidad Católica del Perú. Realiza estudios de Maestría en Psicología Educativa en la U.N.M.S.M.
 3. Bachiller en Psicología. Realiza estudios en el campo de la informática aplicada a la Psicología. Correspondencia enviarla a: Los Olmos 702, Residencial San Felipe, Lima 11, Perú.

La psicometría ha tenido desde sus orígenes con la aparición de la Teoría Clásica de los Tests (TCT), un gran desarrollo caracterizado por la creación de múltiples procedimientos para construir pruebas psicológicas, habiéndose llegado en la actualidad a plantear nuevos métodos para construirlas, destacando principalmente aquellos referidos al estudio de la Estructura Latente del instrumento, como es el caso de la llamada Teoría de la Respuesta al Item (TRI) y en especial del modelo de un parámetro logist de Rasch (Buysen, Van Den Wolleberg y Wimmers, 1983; Muñiz, Rogers y Swaminathan, 1988; Appelbaum 1989).

Avances recientes en el campo de la teorización psicológica han llevado a plantearse la necesidad de evaluar la adecuación tanto conceptual como métrica de las pruebas con las que se trabaja (Marin, 1986), por lo que esto es imprescindible de ser aplicado en todos los ámbitos de acción de la Psicología.

En lo que concierne al campo aplicado de la Evaluación Psicológica, encontramos que el Test de Desarrollo Visomotor de Beery (IVM) ha demostrado ser un instrumento muy útil para evaluar el desarrollo de la integración visomotora (Beery, 1967), motivo por el cual se ha desarrollado en nuestro medio una forma corta del mismo (Salazar y Delgado, 1985).

Es debido a estas consideraciones que nos planteamos la necesidad de llevar a cabo el estudio de las características Psicométricas que presenta la forma corta del test de Beery bajo los supuestos del modelo de Rasch, con miras a tratar de evaluar la adecuación tanto del instrumento como de dicha metodología en nuestro medio.

La forma corta del Test de Beery

La prueba del IVM fue adaptada en I.N.I.D.E. por Salazar y Delgado (1985) y se estudiaron 12 de las 24 figuras geométricas que componen la prueba original (Beery, 1967), habiendo sido seleccionadas sólo aquellas que correspondían a las edades estudiadas.

El instrumento fue administrado en forma exploratoria a un grupo de 245 niños tanto varones como mujeres, cuyas edades se encontraban entre los 6 y 9 años de edad. Se realizó el análisis de ítemes según la metodología utilizada por su autora (Beery, 1967), llegando a establecerse las respectivas tablas de calificación por edades.

Las figuras seleccionadas para la adaptación fueron: (1) El triángulo, (2) El cuadrado y círculo, (3) Las tres líneas cruzadas, (4) Las flechas direccionales, (5) Los aros en 2 dimensiones, (6) El triángulo de 6 círculos, (7) El círculo y cuadrado inclinado, (8) El diamante vertical, (9) Los triángulos inclinados, (10) El círculo de 8 puntos, (11) Los hexágonos de Wertheimer y (12) El diamante horizontal.

Las figuras fueron presentadas en el mismo orden señalado en la secuencia original, pero según los resultados del análisis de ítemes, sería necesario cambiar la secuencia de presentación de las mismas, para que aparezcan en orden de dificultad creciente de acuerdo al desarrollo visomotor que presentan los niños de nuestra realidad.

El Modelo de Rasch

El modelo de Rasch (1966) es un caso especial de los modelos de Estructura Latente, el cual se caracteriza por asumir que cada ítem que mide un rasgo determinado, debe presentar un trazado característico, cuando se grafican las respuestas que dan los sujetos, (Lord y Novick, 1974; Cliff, 1979; Hui, Drasgow y Bei-Hung Chang, 1983; Embretson, 1983; Hulin, 1987; Appelbaum, 1989, Becker, 1990).

Este modelo ha sido relacionado con los procedimientos utilizados para construir las escalas Guttman, en las cuales los ítemes son seleccionados ordenadamente según su nivel de dificultad, de tal manera que cualquier sujeto que falle al ítem X, deberá fallar todos los más difíciles que X y acertar todos los más fáciles que X (Andrich, 1988).

Teóricamente el modelo, asume que cuando el sujeto "v" responde al ítem "i" e indica SI ó NO, la variable respuesta X_{vi} será puntuada como 1 ó 0, dependiendo si la respuesta se califica como de alto (1) o bajo (0) status en el rasgo que está siendo medido. El tipo de respuesta que presente el sujeto a los ítemes, ya sea acierto (1) o error (0), dependerá de:

- (1) La dificultad que presenta el ítem "i" para expresar el rasgo en todos los sujetos, y
- (2) La habilidad o rasgo del sujeto que le permite responder a los ítemes.

Por ejemplo, supongamos que el rasgo del sujeto "v", que viene a ser la medida de la habilidad del sujeto, sea representado por el parámetro β_v (persona-parámetro), y que la dificultad ítem "i" para expresar el rasgo esté representada por el parámetro δ_i (ítem-parámetro).

El modelo de respuesta que permite calcular la probabilidad de la presentación del puntaje X_{vi} correcto (1), tiene una representación matemática (Rasch, 1966; Buysen, Van Den Wollenberg y Wimmers, 1983; Kline, 1986; Van Den Wollenberg, 1982; Andrich, 1988), expresada de la siguiente forma:

$$P(X_{vi}) = \frac{\exp(\beta_v - \delta_i)}{1 + \exp(\beta_v - \delta_i)}$$

Donde el valor más alto de (β_v), será la probabilidad de que el sujeto "v" obtenga un puntaje de 1 para su respuesta y análogamente, el valor más bajo de dificultad (δ_i) del ítem "i", será la probabilidad más alta de que el sujeto tenga un puntaje 1 para esa respuesta.

La propiedad crítica del modelo, radica en que la estimación de la dificultad del ítem (δ_i) es independiente del nivel de habilidad del sujeto (β_v), por lo cual se puede concluir que la dificultad de ítem también será independiente de la distribución del rasgo en la muestra de sujetos cuyas respuestas son analizadas (Kline, 1986; Muñiz, Rogers y Swaminathan, 1988).

En consecuencia, la estimación de la consistencia interna de una escala en una población, puede ser obtenida sin ser contaminada por la falta de varianza en la muestra estudiada. Es debido a ello que la estimación δ_i para cada ítem "i" será equivalente, en independencia de las mediciones realizadas en los sujetos que han sido utilizados como muestra para obtener las respuestas, por lo que se concluye que es factible estudiar la consistencia de la escala entre las poblaciones.

Este modelo tiene 2 finalidades:

- (1) Proporcionar un procedimiento, que permita determinar cuándo una escala es consistente, en independencia de la varianza del rasgo que exista en la población.
- (2) Identificar la interacción que existe entre el ítem y la población, en la medida que la varianza del ítem intrapoblación puede diferir de la varianza del mismo entre poblaciones; si se da el caso que difieren las varianzas,

es posible poner en duda la adecuación del instrumento de medición para efectuar comparaciones entre diferentes poblaciones, como podría ser el caso por ejemplo de los estudios transculturales.

El modelo matemático propuesto por Rasch, presenta 5 axiomas (Rasch, 1966; Buysen, Van Den Wollenberg y Wimmers, 1983; Van Den Wollenberg, 1982), los cuales son:

1. *Dicotomía de los ítemes*.- La variable que se mide, debe ser presentada de forma binaria, ya sea (V o F), acierto o error, aunque es posible generalizar el modelo a ítemes politómicos.

2. *Monotonicidad*.- La probabilidad de la emisión de una respuesta correcta, debe incrementarse a medida que aumentan los valores de los persona-parámetros para todos los ítemes.

3. *Unidimensionalidad*.- Los ítemes debe ser homogéneos, pues se requiere que midan una sola habilidad común a todos ellos, la cual constituye el rasgo latente estudiado.

4. *Independencia local estocástica*.- La única fuente de asociación que debe existir entre los ítemes debe ser el rasgo latente, el cual es fijo, mientras que las respuestas a los ítemes deben ser independientes entre sí.

5. *Suficiencia estadística de la simple suma*.- Toda la información necesaria para calcular los ítemes-parámetros y los persona-parámetros pueden ser computados a partir del número de respuestas correctas que tiene el sujeto.

Cuando todos los supuestos son verificados de forma empírica, se puede afirmar que se cumple la *Objetividad Específica*, la que implica que en el universo de sujetos y medidas, la selección de los ítemes es irrelevante para la medida de los sujetos, y que la selección de los sujetos es irrelevante para la medida de los ítemes. En este caso, la selección aleatoria de los sujetos durante el proceso de construcción de pruebas es irrelevante, aunque la comparación de diversos tipos de muestras de sujetos permite estudiar la validez y generalidad del modelo estudiado. Es debido a todo esto que es posible obtener:

- a. Instrumentos que pueden ser directamente comparables entre sí.
- b. Instrumentos universales construidos a partir de muestras representativas de las poblaciones.
- c. Unidades de medidas que pueden ser catalogadas como de un nivel de intervalo e incluso de razón.

El modelo de Rasch además de ser útil para construir instrumentos de medida, adicionalmente, puede constituirse en un método alternativo para el análisis de datos.

Una de las ventajas que presenta este modelo para construir Pruebas Psicológicas radica en que los ítemes pueden ser usados para obtener los puntajes reales de los sujetos según su verdadero nivel de habilidad, ya que el nivel de dificultad de los ítemes pueden diferenciarse de la habilidad de los sujetos.

Otras ventajas radica en que a diferencia de la teoría clásica de los tests (TCT), si es posible verificar todos sus supuestos de forma empírica (Buyssen, Van Den Wollenberg y Wimmers, 1983).

En la actualidad se han desarrollado numerosos procedimientos para estimar los parámetros δ_i y β_v , así como para identificar aquellos ítemes en los cuales las respuestas no concuerdan con el modelo, la estrategia más recomendable para estimar los parámetros sugerida por los autores, es el método de estimación de la Máxima Verosimilitud Conjunta (Van Den Wollemborg, 1977; Raaijmakers y Van Den Wollemborg, 1979; Gustafsson, 1980; Molenaar, 1983; Traub y Lam, 1985; Rost, 1988).

El cálculo de la estimación de los parámetros, implica la maximización de forma simultánea tanto de la matriz de las probabilidades de respuesta que presentan los ítemes, como de los valores de los parámetros computados.

La prueba de bondad de ajuste al modelo se obtiene analizando si los datos observados directamente pueden ser obtenidos con la aplicación de la fórmula del modelo después que se han estimado los persona-parámetros y los ítemes parámetros.

La evaluación de la consistencia interna de los ítemes en una población es denominado: "forma del ítem intra-población", en tanto que la revisión de la consistencia de los ítemes con el modelo con relación a varias poblaciones es la: "forma del ítem entre-poblaciones.

El procedimiento que se utiliza para realizar estas evaluaciones de la consistencia interna se deriva de la propiedad principal que tiene la estimación de los parámetros, según la cual; "La estimación de los parámetros de los ítemes es independiente de los sujetos que los responden". De tal modo que el parámetro obtenido por un ítem en cada muestra de la población de sujetos estudiados, debe ser comparados estadísticamente para determinar si se obtienen los mismos resultados (Van Den Wollemborg, 1982; Thissen y Steinberg, 1988), y evaluar de ese modo la capacidad de generalización de los resultados.

El modelo de Rasch es matemáticamente mucho más complejo que el modelo clásico de la teoría de los test y su aplicación requiere por lo tanto del uso de un programa estadístico de computación, siendo utilizado para este estudio el programa PML (Molenaar, 1990).

Métodología

La muestra provino de 22 Centros Educativos de los cuales 12 estatales (54%) y 10 no-estatales (46%) ya que se trató de mantener la representatividad de cada tipo de centro educativo en la muestra final, siendo evaluados 1521 niños. El tamaño de la muestra fue calculado a partir de la fórmula de Cochran (1981) para poblaciones con tamaños inferiores a los 100,000 casos. La distribución de la muestra aparece en el Cuadro 1.

Cuadro 1
Distribución compuesta de la muestra

C E N T R O	E D U C	G R A D O	CONTEXTO SOCIO-ECONOMICO				TOTAL	
			URBANO-MARGINAL		URBANO		POR	POR
			MASC.	FEM.	MASC.	FEM.	GRADO	C.E.
E S T A T A L E S	1ro.	80	80	80	80	320	820	
	2do.	70	70	70	70	280		
	3ro.	50	50	60	60	220		
N O	1ro.	21	19	40	40	120	390	
	2do.	30	30	35	35	130		
	3ro.	30	30	40	40	140		
T O T A L	POR SEXO	281	279	325	325	1210		
	POR C.E.	560		650				

Los instrumentos utilizados fueron, la forma corta del test de Beery desarrollada por Salazar y Delgado y una encuesta para obtener los principales datos demográficos de los sujetos.

El Procedimiento de recolección de datos fue como sigue:

- Los datos fueron obtenidos entre mayo y julio de 1987, las aplicaciones fueron grupales, se efectuaron para cada aula en un solo día. Tomando en cuenta el turno de estudio de los sujetos ésta fue realizada sólo en las primeras horas de clases, ya sea por la mañana o por la tarde.
- Para cada evaluación se realizó una fase inicial de motivación y acercamiento a los niños para posibilitar que presentaran su mejor rendimiento. Posteriormente se distribuyeron los protocolos y se dieron las instrucciones pertinentes. Los datos de filiación fueron consignados por los propios niños, excepto en el caso de los del primer grado, en que éstos fueron consignados por los investigadores.
- Cuando los niños iniciaban el desarrollo de la prueba y no deseaban continuarla, se procedió a motivarlos personalmente para que hicieran los dibujos respectivos, de modo que concluyeran la prueba. Durante las aplicaciones estos casos fueron muy poco frecuentes y siempre se logró culminar la aplicación.

Una vez recogidos los datos fueron procesados con el programa PML (Molenaar, 1990) y se utilizó el procedimiento de la Máxima Verosimilitud Conjunta para computar los parámetros del modelo. Adicionalmente se calculó la correlación de Pearson.

Resultados

En el Cuadro 2 apreciamos que la forma corta del VMI no cumple con los requisitos del modelo de Rasch, pues los test estadísticos de Martín Löff y Anderson denotan la existencia de graves distorsiones de los resultados de la escala completa, mientras que al eliminarse los ítemes más fáciles y los más difíciles, y de esa forma conformar una prueba de sólo 6 ítemes, a la que hemos denominado VMI-RASCH, comprobamos que ésta sí cumple con los supuestos del modelo, alcanzando medidas de bondad de ajuste adecuadas.

El estudio de las características psicométricas que presentan los ítemes seleccionados, nos indica que estos conforman una escala de dificultad creciente con correlaciones biseriales puntuales significativas. También encontramos que los niveles de dificultad o ítemes-parámetros (δ) computados, permiten muestrear de manera adecuada los niveles de habilidad medidos en los sujetos.

Cuadro 2
Tests de bondad de ajuste al modelo de RASCH

TEST		VMI	VMI-RASCH
MARTIN LÖF	X ²	159.08	33.17
	GL.	110	20
	p	.002	.132
ANDERSON	X ²	43.86	6.93
	GL.	11	5
	p	.000	.226

Adicionalmente, el estudio de los resultados obtenidos en los test de Molenaar (U₁), confirman la adecuación de cada ítem en la escala.

Cuadro 3
Ítemes-parámetros de la prueba bajo el modelo de RASCH

Nº DE ÍTEM	p	r _{pbis}	δ	MOLENAAR U ₁
4	0.58	0.74	-1.98	0.071
5	0.50	0.70	-0.93	0.673
6	0.45	0.70	-0.33	2.297
8	0.33	0.76	0.30	-2.062
9	0.16	0.69	1.41	-1.671
10	0.14	0.69	1.53	-0.807

El análisis de los Personas-Parámetros, nos indica que la habilidad estudiada se presenta de manera diferenciada en la muestra de sujetos y que tanto el índice de separación como el coeficiente de Kuder-Richardson₂₀ nos permiten establecer que la puntuación asignada es consistente.

El estudio de la capacidad de generalización del modelo, en el cual se evalúa hasta qué punto la prueba mantiene la forma del modelo considerando la división de la muestra en diferentes variables, fue realizado a través del test de Anderson, el cual nos indica que ni la edad, el grado o el sexo, son elementos que produzcan sesgo alguno que invalide la adecuación del modelo, con lo cual comprobamos la constancia de la forma del instrumento en toda la muestra.

Cuadro 4
Personas-parámetros de la prueba bajo el modelo de RASCH

PUNTAJE	B	
1	-2.09	
2	-0.92	
3	0.02	
4	0.95	
5	2.07	
INDICE DE SEPARACION		0.60
KUDER RICHARDSON 20		0.70

Cuadro 5
Test de Anderson de la razón de verosimilitud de la igualdad de los ítems-parámetros por grupos estudiados

VARIABLES	Nº DE GRUPOS	X ²	GL.	p
EDAD	3	11.88	10	.293
GRADO	3	13.18	10	.214
SEXO	2	1.71	3	.635

Finalmente, encontramos que existe una elevada relación entre el VMI y el VMI-RASCH ($r = 0.98$), dato que concuerda con los hallazgos reportados por diversos autores al relacionar una escala con su correspondiente versión RASCH (Andrich, 1988).

Conclusiones

Los resultados obtenidos nos permiten llegar a las siguientes conclusiones:

- a) Los ítems que conforman la versión corta del Test de Beery (VMI) no constituyen una escala que tenga las propiedades del modelo de RASCH.
- b) Los ítems 4, 5, 6, 8, 9 y 10 conforman una escala que cumple con todas las propiedades del modelo estudiado, y que será denominada VMI-RASCH.
- c) El VMI-RASCH es un instrumento que tiene la propiedad de especificidad objetiva, y por lo tanto sus resultados son independientes de la muestra utilizada para su construcción.

Referencias

- Andrich, D. (1988). *Rasch models for measurement*. Londres: Sage Publications Ltd.
- Appelbaum, J. (1989). Psychometric methods. *Annual Review of Psychology*, 40, pp. 23-43.
- Becker, B. (1990). Item characteristics and gender differences on the SAT1-M for mathematically able youths. *American Educational Research Journal*, 27, 1, pp. 65-87.
- Beery, Keith (1967). *Visual-Motor Integration Monograph*. Chicago: Follet Publishing Company.
- Buysse, H.; Van Den Wollenberg A. y Wimmers M. (1983). *Construction of a Self-Esteem Scale The Latent Trait Alternative*. Informe inédito, K.U. Nijmegen, Holanda.
- Cliff, N. (1979). La psicometría. En Wolman B. (Ed.) *Manual de Psicología General: Historia, teoría y método*. Barcelona, Ed. Martínez Roca S.A., 1.
- Cochran, W. (1981). *Técnicas de Muestreo*. México: Compañía Editora Continental S.A.
- Embretson S. (1983). Construct Validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 1, pp. 179-197.
- Gustafsson, J. (1980). A solution of the conditional estimation. Problem for long test in the Rasch model for dichotomous items. *Educational and Psychological measurement*, 40, pp. 377-385.
- Hui, H.; Drasgow, F. y Bei-Hung Chang (1983). Analysis of the modernity scale: An item response theory approach. *Journal of Cross Cultural Psychology*, 14, 3, pp. 259-278.
- Hulin, Ch. (1987). A psychometric theory of evaluations of item and scale translations: fidelity across languages. *Journal of Cross Cultural Psychology*, 18, 2, pp. 115-142..
- Kline, P. (1986). *A Handbook of test construction: introduction to psychometric design*. Londres: Methuen y Co. Ltd.
- Lord, F. y Novick, M. (1974). *Statistical theories of mental test scores*. Londres, Addison-Wesley Publishing Company.
- Marin, G. (1986). Consideraciones metodológicas básicas para conducir investigaciones psicológicas en América Latina. *Acta Psiquiátrica y Psicológica de América Latina*, pp. 183-192.
- Muñiz, J. Rogers, J. y Swaminathan, H. (1988). Robustez de las estimaciones del Modelo de Rasch en presencia de aciertos al azar y discriminación variable de los ítemes. *Anuario de Psicología*, 43.
- Molenaar, I.W. (1983). Some improved diagnostics for failure of the Rasch model. *Psychometric*, 48, 1, pp. 49-71.
- Molenaar, I.W. (1990). *PML: users manual, PC version*. IEC proGAMMA, Groningen, Holanda.

- Raaijmakers, M.H. y Van Den Wollenberg, A. (1979). *RADI: Programme for the dichotomus Rasch model*. Informe inédito, K.U. Nijmegen, Holanda.
- Rasch, G. (1966). An item analysis which takes individual differences into account. *The British Journal of Mathematical and Statistical Psychology*, 19, Part. 1, pp. 49-57.
- Rost, J. (1981). Measuring attitudes with a threshold model drawing on an traditional scaling concept. *Applied Psychological Measurement*, 12, 4, pp. 397-407.
- Salazar, V. y Delgado, Ana (1985). *Adaptación de la Prueba de Desarrollo de Integración Visomotriz*. Lima: INIDE.
- Thissen, D. y Steinberg, L. (1988). Data Analysis using item response theory. *Psychological Bulletin*, 104, 3, pp. 385-395.
- Traub, R. y Lam, R. (1985). Latent structure and item sampling models for testing. *Annual Review of Psychology*, 36, pp. 19-48.
- Van Den Wollenberg (1977). *Some structural models for the pencil and cylinder test*. Informe inédito, K.U. Nijmegen, Holanda.
- Van Den Wollenberg (1982). A simple and effective method to test the dimensionality axiom of the Rasch model. *Applied Psychological Measurement*, 6, 1, pp. 83-91.
- Weiss, D. y Davidson, M. (1981) Test theory and methods. *Annual Review of Psychology*, 32, pp. 629-658.