

UNA REVISION DE LAS PROPIEDADES PSICOMETRICAS DEL TEST DE MATRICES PROGRESIVAS DE RAVEN (ESCALA SUPERIOR)

Esther TORRES, y Marcelino CUESTA

Euskal Herriko Unibertsitatea
Universidad de Oviedo

RESUMEN

Este trabajo presenta una revisión de las propiedades psicométricas del Test de Matrices Progresivas de Raven (Escala Superior), empleando para ello una muestra de 1028 sujetos estudiantes de primer curso de la Universidad del País Vasco/Euskal Herriko Unibertsitatea. Los análisis realizados se llevan a cabo tanto desde la perspectiva clásica (fiabilidad, dificultad, discriminación) como desde la Teoría de Respuesta a los Items (función de información, curva característica del test parámetros a y b). Según los resultados obtenidos, cabe destacar la facilidad de la prueba para el tipo de sujetos empleados y la poca información que aporta para niveles medios y altos de la variable medida.

Palabras Clave: Test de Raven, Teoría de Respuesta a los Items, Modelos Logísticos, Función de Información.

ABSTRACT

Psychometrica properties of the Raven progressive matrices test. This paper presents a review of the psychometric properties of Raven's Progressive Matrices test (superior level) using a sample of 1028 freshman from the University of País Vasco/Euskal Herriko Unibertsitatea. The analyses were carried out from the Classical Test Theory perspective (reliability, difficulty, discrimination) and from the Item Response Theory point of view (information function, test characteristic curve, parameter a and b). The results show the easiness of this test for the kind of subjects used here, and the poor information provided for middle and superior values of the Θ scale.

Key words: Raven Test, Item Response Theory, Logistic Models, Information Function.

En este trabajo se lleva a cabo un análisis psicométrico de uno de los tests más clásicos y de más amplia utilización en la medida de la Inteligencia General, el Test de Matrices Progresivas de Raven. Si bien las características psicométricas de esta prueba están bien documentadas, lo "veterano" de la edición española así como los resultados encontrados por Cuesta, Muñiz y Paz (en prensa) al emplear la Teoría de Respuesta a los Items (TRI) en un trabajo similar al que aquí se presenta, hacen pensar en lo conveniente de llevar cabo una revisión de sus propiedades. El presente trabajo se centra principalmente en el análisis de items aplicando para ello los modelos de TRI. El objetivo es conocer los índices de dificultad, discriminación y la función de información de los items, además de descubrir como cada uno de los items del Raven contribuyen a la medida de la Inteligencia General. Para ello contamos con una muestra amplia de 1028 sujetos, estudiantes de la Universidad del País Vasco/E.H.U., de manera que podamos tener una cierta seguridad en la estimación de los parámetros, y de forma especial el parámetro de discriminación. (Baker, 1987; Lord, 1986).

Los análisis se realizan desde un punto de vista clásico (Gulliksen, 1950; Lord y Novick, 1968; Crocker y Algina, 1986) y también desde el punto de vista de la TRI, que como es bien sabido aporta ciertas alternativas a algunos problemas que la Teoría Clásica de los Tests (TCT) no podía abordar (Lord, 1980; Hambleton y Swaminathan, 1985; Goldstein y Wood, 1989; Muñiz, 1990; Muñiz y Hambleton, en prensa).

METODO

Muestra

Participaron 1.028 sujetos de los cuales 542 eran hombres y 486 mujeres. Sus edades estaban comprendidas entre 17 y 42 años, con una media de 18,95 años. Todos los suje-

tos eran estudiantes de primer curso universitario pertenecientes al Campus de Gipuzkoa de la Universidad del País Vasco/Euskal Herriko Unibertsitatea durante el curso académico 1989/1990. La distribución de la muestra por centros fue como sigue: 152 sujetos de la Facultad de Derecho, 232 de la Escuela Técnica Superior de Ingenieros Industriales, 251 de la Facultad de Informática, 132 de la Facultad de Psicología, 40 de la Facultad de Filosofía y Ciencias de la Educación, 78 de la Escuela de E.G.B. y 143 de la Escuela Universitaria de Estudios Empresariales.

Instrumento

Test de Matrices Progresivas de Raven, versión de 1962 y editado por Mepsa en 1970. Esta versión consta de dos series, la serie I compuesta por 12 items y la serie II por 36.

Procedimiento

La aplicación del test se realizó durante el primer trimestre de 1990 en los respectivos centros, necesiándose varios días para finalizar la administración. La participación de los sujetos fue voluntaria. La prueba se realizó siguiendo las instrucciones del manual y respetando el tiempo de aplicación propuesto en el manual (40 minutos).

Análisis de Datos

En la realización de los análisis se emplearon dos paquetes de ordenador: SPSS/PC+, para los análisis clásicos, y PC-BILOG (Mislevy y Bock, 1986), para los análisis específicos de TRI.

Razones teóricas hacían suponer que dadas las características de la prueba que nos ocupa el modelo logístico más apropiado sería el de dos parámetros, pues con seis alternativas de respuesta por item parece razonable esperar que el azar juegue un papel

poco importante y por tanto $c=0$. Por otro lado los índices de discriminación clásicos diferentes permiten desechar el modelo de un parámetro. Efectivamente, tras comprobación empírica del ajuste de los modelos de uno, dos y tres parámetros, el modelo logístico que mejor ajusta a nuestros datos es el de dos parámetros ($\chi^2 = 266.9$, $p \leq 0.6728$). Sólo dos de los cuarenta y ocho ítems que componen el test no ajustan a dicho modelo, son los ítems 11 y 12.

RESULTADOS

Índices de dificultad y discriminación de los ítems

En la Tabla 1 se pueden observar los índices de dificultad y de discriminación tanto desde el punto de vista de la TCT como desde la TRI. Se observa la coincidencia de ambos tipos de índices en cuanto a las conclusiones a las que apuntan, el test resulta muy fácil y poco discriminativo para el tipo de sujetos que componen la muestra en la que trabajamos.

Ahondando un poco más en lo que acabamos de apuntar más arriba, señalar que la media del parámetro b (dificultad) resultó ser -1.798 , dándose casos extremos de ítems como el 1 y 2 con valores para b de -7.583 y -5.083 (nótese que la escala de Θ suele representarse a lo sumo entre $+4$ y -4).

En lo que respecta al parámetro a, recordando que su rango de valores suele situarse entre 0 y 2, tomamos como referencia, aún manteniendo todas nuestras reservas hacia las etiquetaciones rígidas, la asignación realizada por Baker (1985, pag 24) a los valores de a.

Como puede verse nuestros valores del parámetro a caerían en el mejor de los casos en la categoría de moderados. Así mismo cabe destacar como los peores índices de discriminación predominan entre aquellos ítems cuya dificultad es algo superior, lo cual

| ITEM | I.DIF. | b | I.DISC. | a | I(Θ)* |
|------|--------|--------|---------|-------|-------|
| 1 | .998 | -7.583 | -.048 | .514 | .1911 |
| 2 | .994 | -5.083 | .086 | .671 | .3249 |
| 3 | .974 | -4.158 | .139 | .574 | .2376 |
| 4 | .990 | -3.445 | .188 | 1.052 | .7988 |
| 5 | .964 | -3.676 | .185 | .601 | .2607 |
| 6 | .898 | -3.687 | .152 | .371 | .0997 |
| 7 | .965 | -3.392 | .212 | .677 | .3314 |
| 8 | .894 | -1.976 | .328 | .827 | .4940 |
| 9 | .756 | -1.609 | .265 | .467 | .1573 |
| 10 | .918 | -2.700 | .232 | .619 | .2768 |
| 11 | .603 | -.464 | .323 | .672 | .3258 |
| 12 | .811 | -1.459 | .309 | .756 | .4132 |
| 13 | .949 | -2.489 | .308 | .916 | .6064 |
| 14 | .969 | -2.866 | .273 | .918 | .6095 |
| 15 | .947 | -2.369 | .343 | .978 | .6910 |
| 16 | .945 | -2.211 | .343 | 1.078 | .8391 |
| 17 | .927 | -2.295 | .337 | .844 | .5144 |
| 18 | .969 | -3.097 | .231 | .810 | .4742 |
| 19 | .912 | -2.181 | .336 | .811 | .4758 |
| 20 | .895 | -2.262 | .282 | .676 | .3306 |
| 21 | .963 | -2.654 | .263 | .969 | .7688 |
| 22 | .901 | -1.783 | .367 | 1.053 | .8016 |
| 23 | .942 | -2.207 | .349 | 1.045 | .7891 |
| 24 | .925 | -2.133 | .319 | .943 | .6424 |
| 25 | .772 | -1.622 | .270 | .509 | .1875 |
| 26 | .884 | -2.388 | .254 | .585 | .2473 |
| 27 | .791 | -1.337 | .361 | .750 | .4065 |
| 28 | .834 | -1.688 | .327 | .700 | .3539 |
| 29 | .793 | -2.207 | .185 | .390 | .1097 |
| *30 | .674 | -.967 | .249 | .509 | .1872 |
| 31 | .809 | -1.995 | .249 | .477 | .1641 |
| 32 | .780 | -2.917 | .119 | .265 | .0509 |
| *34 | .653 | -.789 | .260 | .532 | .2047 |
| *35 | .726 | -1.015 | .343 | .674 | .3278 |
| 33 | .748 | -1.196 | .316 | .646 | .3014 |
| *36 | .460 | .261 | .248 | .474 | .1620 |
| *37 | .689 | -.988 | .309 | .522 | .1972 |
| *38 | .594 | -.764 | .174 | .285 | .0585 |
| *39 | .564 | -.267 | .327 | .489 | .1728 |
| *40 | .431 | .558 | .265 | .408 | .1201 |
| 41 | .328 | 1.115 | .258 | .472 | .1607 |
| *42 | .592 | -.528 | .275 | .383 | .1057 |
| *43 | .557 | -.191 | .349 | .550 | .2184 |
| *44 | .491 | .077 | .206 | .335 | .0809 |
| *45 | .492 | .085 | .143 | .266 | .0510 |
| *46 | .440 | .382 | .292 | .481 | .1670 |
| 47 | .840 | -2.680 | .162 | .395 | .1127 |
| 48 | .163 | 2.537 | .154 | .415 | .1243 |

* Los puntos de Máxima Información en el modelo logístico de dos parámetros coinciden con el valor de b.

Tabla 1: Índices de Dificultad, Discriminación e Información

viene a reafirmar lo arriba dicho sobre la facilidad del test para nuestra muestra.

| Etiquetas Verbales | Valores Numéricos |
|---------------------------|--------------------------|
| Ninguna | 0 |
| Muy baja | 0,01–0,34 |
| Baja | 0,35–0,64 |
| Moderada | 0,65–1,34 |
| Alta | 1,35–1,69 |
| Muy alta | > 1,70 |
| Perfecta | + infinito |

Fiabilidad y Dimensionalidad de la escala

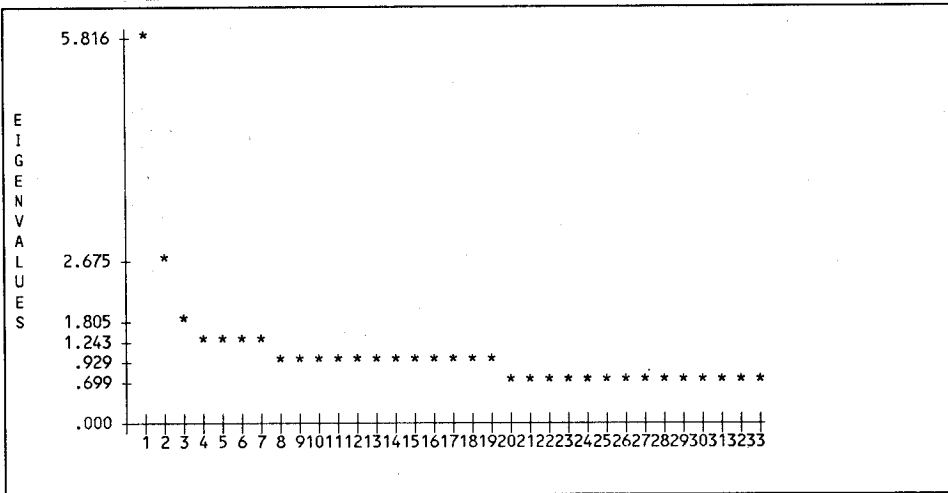
La fiabilidad fue estimada mediante el coeficiente Alfa (α), tomando un valor de 0,83. Como complemento a este índice de consistencia interna se calcularon otros dos indicadores derivados del análisis factorial como son la Theta (Θ) de Carmines (Carmines y Zeller, 1979) y la Omega (Ω) de Heise y Bohrnstedt (1970). Como es conocido estos

coeficientes son estimadores del límite superior de α siendo su ordenación en lo referente a su cuantía como sigue $\alpha < \Theta < \Omega$. Para los datos con los que hemos trabajado Θ alcanza un valor de 0,845 y ω de 0,91.

Tomando como criterio el buen ajuste del modelos logístico de dos parámetros a los datos ($\chi^2 = 266,9, p \leq 0,6728$), puede considerarse que los items constituyen un constructo razonablemente unidimensional. Sin embargo, sometidos los items a un análisis factorial (PCA), la unidimensionalidad aparece como más problemática (véase gráfica 1). Ambos criterios parecen no converger estrictamente, pero son bien conocidas las limitaciones del análisis factorial con variables dicotómicas (McDonald, 1981; Hattie, 1984; Bernstein y Teng, 1989).

Curva característica del test

Cuando se trabaja en el marco de la TRI las puntuaciones de los sujetos vienen dadas en lo que conocemos como escala Θ , cuyos límites teóricos están entre $+\infty$ y $-\infty$, aunque a nivel práctico los valores entre los cuales se



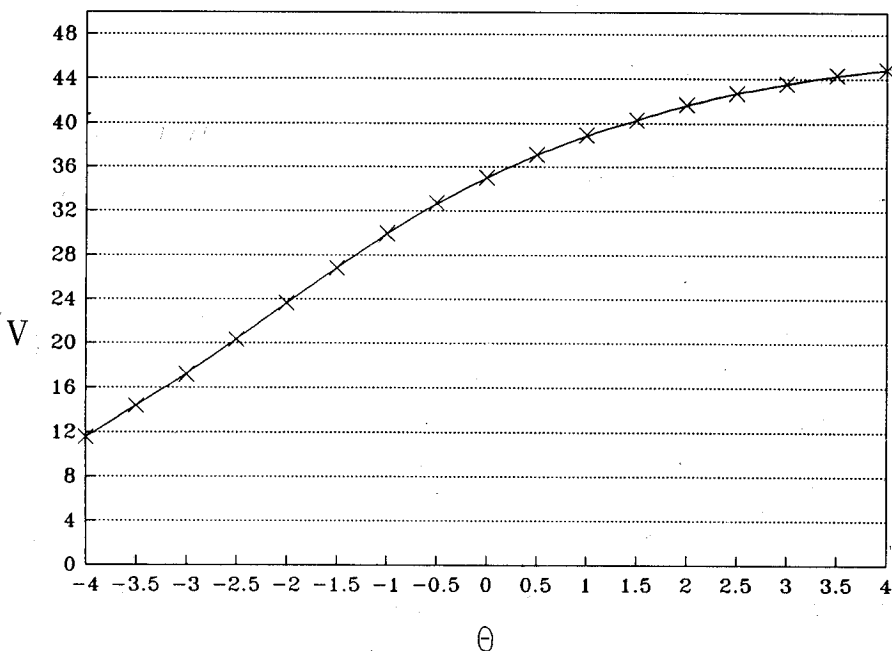
Gráfica 1: Scree-plot del análisis factorial de los items.

suele tomar van entre +3 y -3. Una manera de establecer relación entre la escala de Θ y la escala propia del test es a través de lo que conocemos como Curva Característica del Test, la cual permite la traducción de un tipo de puntuación a otra. Estas dos puntuaciones son una misma cosa pero expresada en diferentes escalas (Lord, 1980; Muñiz, 1990). La curva que expresa esta relación en nuestros datos puede observarse en la siguiente gráfica.

En el marco de la TRI la Función de Información de un Test, $I(\Theta)$, viene dada por:

$$I(\theta) = \sum_{i=1}^n \frac{[P'_i(\theta)]^2}{P_i(\theta) Q_i(\theta)} \quad (2)$$

Y nos indica la información que el test aporta para los diferentes niveles de Θ ,



Gráfica 2: Curva Característica del Test según el modelo logístico de dos parámetros.

Función de información

En general cuando nos referimos al concepto de Información en el ámbito de la estimación de parámetros, este viene definido por el inverso de la varianza de error (De Gruitjer y Van der Kamp, 1984; Hulin, Drasgow y Parsons, 1983):

$$I = \frac{1}{\sigma_e^2} \quad (1)$$

cuanta más información más precisión en la medida o lo que es lo mismo menos error de medida.

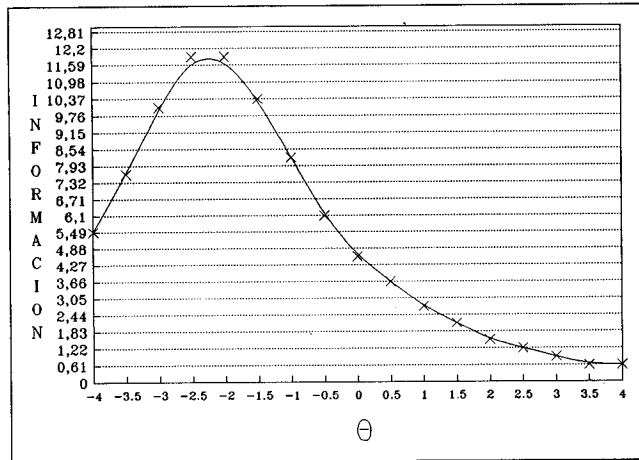
Como se ve en la expresión 2 la Función de Información del Test es la suma de las Funciones de Información de los Items. Fijándonos en los puntos de máxima información que aparecen en la tabla 1 observamos como sólo 14 (los señalados con un asterisco) de los 48 items aportan información máxima en valores medios de la escala (entre

± 1). Sin embargo, en esos puntos la cantidad de información es mínima, dándose el mayor aporte de información en los ítems situados en niveles más bajos de Θ .

En lo referente al test en su conjunto, en la gráfica 3 ofrecemos su Función de Información.

En concordancia con todo lo que llevamos expuesto nos encontramos con que la mayor efectividad del test que nos ocupa se da para valores de Θ aproximadamente entre -3 y -1 , luego su ámbito ideal de utilización será con sujetos cuya puntuación en Θ se sitúe en esa franja.

Nótese que la utilización de técnicas procedentes del campo de la TRI no viene más que a precisar más la información que aportan los índices clásicos. Gran parte del "arsenal" tecnológico que aportan estos modelos ha sido aquí obviado (como la Función de Información de cada ítem, o la Curva Característica de los Ítems, que también nos ofrece PC-BILOG) por cuanto supondría sobrepasar los límites de lo que este trabajo pretendía. Es claro, en cualquier caso, que en la línea de otros trabajos (Cuesta et al., en prensa) también desde aquí animamos a las empresas españolas del ramo a esforzarse en



Gráfica 3: Función de Información del Test según el modelo logístico de dos parámetros.

CONCLUSIONES

A la luz de los datos que se han ido ofreciendo la conclusión más general es que el Test de Matrices Progresivas de Raven (Escala superior) aporta su máxima información para valores muy bajos de la variable inteligencia. Parece, por tanto, desaconsejable la utilización de este instrumento cuando el deseo del profesional sea discriminar entre sujetos de alta competencia.

la utilización del nuevo instrumental, no sólo en la actualización de tests existentes sino, y más bien, en la creación de nuevas pruebas dentro de la filosofía de los bancos de ítems donde la TRI despliega realmente todo su potencial, tal y como resaltan Muñiz y Hambleton (en prensa): "Los Bancos de Ítems constituyen la piedra angular de la TRI, sin ellos nada de lo dicho hasta ahora se mantendría y sería aplicable".

REFERENCIAS

- Baker, F.B. (1985). *The basics of item response theory*. Porstmouth: N. H. Heineman.
- Baker, F.B. (1987). Item parameter estimation under one, two and three parameter logistic models. *Applied Psychological Measurement*, 11, 11-141.
- Bernstein, I.H. y Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, 105, 467-477.
- Carmines, E.G. y Zeller, R.A. (1979). *Reliability and validity assessment*. Londres: Sage.
- Crocker, L. y Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Cuesta, M.; Muñiz, J. y Paz, M.D. (en prensa). Funciones de información y otras propiedades psicométricas de cuatro tests clásicos de inteligencia general: PMA-R, Raven, D-48 y BLS-IV. *Investigaciones Psicológicas*.
- De Gruitjer, N.M. y Van der Kamp, L.J.T. (1984). *Statistical models in psychology and educational testing*. Lisse: Swets and Zeitlinger.
- Goldstein, H. y Wood, R. (1989). Five decades of item response modelling. *British Journal of Mathematical and Statistical Psychology*, 42, 139-167.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Hambleton, R.K. y Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer Academic Publishers.
- Hattie, J.A. (1984). An empirical study of various indices for determining unidimensionality. *Multidimensional Behavioral Research*, 19, 49-78.
- Hattie, J.A. (1985). Methodological review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.
- Heise, D.R. y Bohrnstedt, G.W. (1970). Validity, invalidity and reliability. En E.F. Borgatten y G.W. Bohrnstedt (Eds.): *Sociological Methodology*. San Francisco: Jossey Bass.
- Hulin, C.L.; Drasgow, F. y Parsons, C.K. (1983). *Item response theory. Application to psychological measurement*. Homewood, Illinois: Dow Jones-Irvin.
- Lord, F.M. y Novick, M. (1968). *Statistical theories of mental tests scores*. Reading, MA: Addison-Wesley.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: LEA.
- Lord, F.M. (1986). Maximum likelihood and bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, 23(2), 157-162.
- McDonald, R.P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.
- Mepsa (1970). *Test de matrices progresivas de Raven. Escala Superior*. Madrid.
- Mislevy, R.J. y Bock, R.D. (1986). *PC-BILOG*. Moresville, IN: Scientific Software, Inc.
- Muñiz, J. (1990). *Teoría de respuesta a los ítems*. Madrid: Pirámide.
- Muñiz, J. y Hambleton, R.K. (en prensa). Medio siglo de teoría de respuesta a los ítems. *Anuario de Psicología*.