

## OBJETIVIDAD Y CONFIABILIDAD DE LA PRUEBA GOODENOUGH-HARRIS EN COSTA RICA

*Heidy Ceciliano Rogers  
Sharon S. Woodburn*

### Introducción

Los esfuerzos para desarrollar técnicas de medición de la inteligencia comenzaron a fines del Siglo XIX (Lindeman, 1971). Según diferentes autores (Lindeman, 1971; Mehrens y Lehmann, 1982; Davidoff, 1984) en la actualidad hay un centenar de pruebas de inteligencia que los educadores en otros países aprovechan, entre ellas la "WAIS" (Escala de Wechsler para Medir la Inteligencia en Adultos) (Wechsler, 1949) la adaptación de esta prueba para niños entre 5 y 15 años (Wechsler, 1974), la "WPPSI" (Wechsler, 1949); "McCarthy Scale of Children's Ability" (Índices de McCarthy para Medir Capacidades en Niños) (McCarthy, 1972); la Escala de Inteligencia Infantil (Cattell, 1960) y la Prueba de Goodenough-Harris (PGH) (Harris, 1963).

Szekely (1960), considera que se puede deducir la inteligencia del niño, partiendo de la forma en que éste hace la representación gráfica de la figura humana, mediante el número de elementos que incluye en ella. El pionero en el desarrollo de este tipo de instrumento fue Florence Goodenough (Szekely, 1960), quien publicó su prueba en 1926. En ella se le pedía que dibujara de la mejor manera posible una persona; cuando terminaba, se le solicitaba que la dibujara de nuevo, a ser posible, mejor que la primera vez. La calificación se basaba en 31 elementos, cada uno con un valor de un punto.

La Prueba de Goodenough fue reformulada por D.B. Harris, (1963) y se denominó a partir de este entonces la Prueba de Goodenough-Harris. En su trabajo presentó la prueba estandarizada en una muestra de 2975 niños de 5 a 15 años de edad en la cual hubo cerca de 75 niños en cada uno de los once niveles de edad (Naglieri y Maxwell, 1981). Harris (1963) informó que el coeficiente de objetividad (o sea, la confiabilidad entre evaluadores) fue de 0,92 y 0,98. También recomendó que se empleara el promedio de las calificaciones

de la figura masculina y la femenina con el fin de lograr una puntuación final válida.

Se han realizado estudios con el objeto de comparar la PGH con diferentes pruebas de inteligencia: la Stanford-Binet (Terman, 1970), "Culture Fair Intelligence Test" (Prueba de Inteligencia, Libre de Efectos Culturales) (Cattell, 1983) y "McCarthy Scale of Children's Ability" (Índices de McCarthy para Medir Capacidades en Niños) (McCarthy, 1972). Lewis y Livson (1977) hicieron un estudio con 106 niños (51 varones y 55 niñas de 5 a 11 años) en que se correlacionaron la PGH y la Stanford-Binet (SB) con el fin de explorar las discrepancias entre las dos pruebas debido a las diferentes características posibles de la personalidad. Los resultados arrojaron correlaciones de 0,41 en varones y 0,42 en niñas, las cuales según los autores están dentro de lo normal.

Weis (1980), comparó los resultados de niños peruanos en la PGH con los de "Culture Fair Intelligence Test" (Cattell, 1963), la cual es una prueba no verbal de inteligencia fluida que se desarrolló con el fin de minimizar la influencia de factores socio-económicos, educación y experiencia cultural. El propósito de su estudio fue determinar si las escalas de clasificación de estas pruebas servían en una cultura diferente a la que se estandarizó. Al aplicar las dos pruebas, se encontró que las puntuaciones en la PGH se ubicaron dentro de las normas estandarizadas en Estados Unidos, mientras que los resultados de la "Culture Free Intelligence Test" estuvieron por debajo de las normas. Después, Weis (1981) utilizó la PGH para medir el CI en dos poblaciones de niños peruanos, una semi-urbana y una rural, cuyas características culturales fueron muy diferentes. Los resultados demostraron que la PGH no era completamente libre de efectos culturales, y por lo tanto, el autor recomienda que los puntos en la escala deben ser estandarizados en cada grupo cultural, ya que puede tener una influencia profunda en los resultados de esta prueba.



Otro estudio de correlación fue llevado a cabo por Naglieri y Maxwell (1981) quienes lograron correlacionar significativa y positivamente la PGH con la sección Dibujar un Niño de "McCarthy Scale of Children's Abilities" (McCarthy, 1972), la cual consiste en sólo 10 ítemes (en vez de 73 y 71 en la PGH). Fue estandarizado en 1972 (Naglieri y Maxwell (1981) con una muestra de 1032 niños de 2 1/2 a 8 1/2 años. El propósito de este estudio fue explorar la diferencia entre los sistemas de evaluación de las dos pruebas en cuanto a la objetividad de los evaluadores con una muestra de 60 niños; resultaron correlaciones mayores de 0,80.

Golomb (1977) efectuó un estudio con cinco pruebas relacionadas con la figura humana, una de las cuales era dibujar una persona; en ella se utilizó la escala de la Prueba Goodenough-Harris para la calificación. Uno de los propósitos de esta investigación fue estudiar los efectos del nivel socio-económico, cociente intelectual, sexo y edad en la repre gráfica de la figura humana. Se concluyó que ninguna de las cinco tareas se ve influenciada por el status socio-económico del niño. Sin embargo, en todas las correlaciones entre el CI y la edad se observó una relación estadísticamente significativa al nivel de 0,001, mientras que se notó una relación significativa de 0,01 entre CI y sexo.

Evaluar una prueba científicamente es importante para definir si puede ser utilizada para determinado fin. Cuando se pretende usarla como instrumento de diagnóstico, es requisito imprescindible que ésta sea válida, objetiva y confiable para la población a la cual se va a aplicar. Una prueba es válida si mide en realidad la conducta específica que se somete a comprobación (Lindeman, 1971). La objetividad y confiabilidad (estabilidad) de la prueba se refieren a la consistencia de la prueba en dos aplicaciones, tanto con evaluadores diferentes como con el mismo evaluador. La confiabilidad interna es la relación que existe entre los elementos evaluados y se calcula con los datos obtenidos de una sola aplicación (Schwanda, 1971).

El presente estudio pretende dar un aporte a la educación escolar, al someter a un análisis de objetividad y confiabilidad la Prueba Goodenough-Harris en nuestro medio para determinar la madurez intelectual en niños de 3 a 15 años de edad ya que uno de los problemas que enfrenta la educación escolar nacional es la carencia de pruebas científicas adaptadas o apro-

piadas para nuestro medio. Se conocen en el país solamente cuatro pruebas que cumplen con los requisitos científicos de validez, confiabilidad y objetividad: La Prueba de Madurez para la Lectura (Méndez y Stover, 1980); el "Test" de la Escuela Meeting Street (Woodburn y Méndez, sin publicar) y la Escala KTRS (Rodríguez-Barquero, 1985).

## Métodos y materiales

### Muestra

Se seleccionaron dos escuelas de la provincia de Heredia: La Escuela Villalobos de Lagunilla y la Escuela Los Lagos, ambas de población semiurbana. En la primera se trabajó con 49 niños (12 varones y 14 niñas de quinto y 16 varones y 7 niñas de sexto grado) y en la segunda participaron 34 (4 varones y 13 niñas de quinto y 2 varones y 15 niñas de sexto grado); sus edades oscilaron entre 10 y 15 años. Para lograr estudiar cada detalle de la prueba, se estimó que a estas edades podrían hacer dibujos que reunieran los ítemes descritos en el manual. La única condición que se requirió en la selección de los niños fue encontrarse en quinto o sexto grado independiente de la edad.

### Instrumento

La Prueba de Goodenough-Harris mide conceptos abstractos, espaciales y cuantitativos, con el fin de determinar la madurez intelectual, o cociente intelectual, en niños de 3 a 15 años de edad y consiste en hacer dos representaciones gráficas de la figura humana, una masculina y otra femenina. Se evalúa la presencia de diferentes elementos (cabeza, ojos, nariz, boca, manos, dedos de la mano, hombros, tronco, y otros) y sus detalles, cada uno de los cuales tiene un valor de un punto; en el caso del dibujo de un hombre hay 73 elementos y 71 para el de una mujer.

### Diseño y Procedimiento

El presente fue un estudio diagnóstico de tipo ex-post-facto que se realizó en grupo (Ceciliano-R, 1988). Las indicaciones fueron dadas al inicio y



no se permitieron aclaraciones intermedias en el transcurso de la misma. No hubo límite de tiempo para terminar la tarea. A cada niño se le suministró un lápiz y un formulario. En la primera hoja él anotó sus datos personales. Se le pidió que dibujara un hombre de la mejor manera posible; cuando terminó, se le pidió que dibujara una mujer. Posteriormente se calificó cada ítem empleando los criterios explicados en el manual. Al establecer la objetividad, la confiabilidad y el error estándar de la prueba se emplearon los lineamientos de Popham (1981). Para determinar la objetividad, el instrumento fue aplicado en la Escuela Villalobos de Lagunilla por un evaluador y ocho días después otro evaluador lo aplicó nuevamente a los mismos niños. Cada uno calificó por aparte las pruebas de los niños con que trabajó. Posteriormente, con los resultados de ambas aplicaciones, se calculó la correlación de Pearson (Popham, 1981). En el procedimiento para determinar la confiabilidad (estabilidad) del instrumento, un mismo evaluador lo pasó dos veces en la Escuela Los Lagos con ocho días entre aplicaciones. Con los resultados de ambas aplicaciones se calculó la correlación de Pearson. Para determinar la confiabilidad interna de la PGH se tomaron los resultados de los participantes de las dos escuelas ( $n = 83$ ) y se aplicaron la fórmula Spearman-Brown (Popham, 1981) y la Kuder-Richardson, (Popham, 1981) tanto con los datos de la primera aplicación, como de la segunda. El error estándar de la prueba fue calculado con los resultados de las dos muestras, tanto para la primera aplicación como la segunda.

## Resultados y análisis

En el Cuadro 1 se pueden observar los promedios y desviaciones estándar de los resultados empleados para obtener la objetividad de la PGH ( $n = 49$ ). Se puede notar que en estos seis grupos de datos la variabilidad de los elementos tienden a disminuir en el dibujo de una mujer; es decir, la desviación estándar en el dibujo de la mujer es menor que en el de un hombre, además se comprobó, mediante la aplicación de la Prueba T de Student (Thomas y Nelson, 1985), que no hubo diferencias estadísticamente significativas a nivel de 5% entre los puntos alcanzados en el dibujo de un hombre y de una mujer, ni tampoco entre los resultados de las dos aplicaciones (ver Cuadro 2).

El Cuadro 3 presenta los resultados de la objetividad, confiabilidad (estabilidad) y confiabilidad interna de la prueba según la correlación de Pearson, la fórmula de Spearman-Brown y el coeficiente de Kuder-Richardson. En cuanto a la objetividad y la confiabilidad (estabilidad) se nota que todas las correlaciones son mayores de 0,78. Se considera que esto se debe a que el manual instructivo incluye indicaciones claras y detalladas, y por tal motivo se evitan interpretaciones diferentes cada vez que se aplica la prueba.

Al calcular la confiabilidad interna del instrumento en la primera aplicación, según la fórmula de Spearman-Brown, el resultado para el dibujo de un hombre fue de 0,91 ( $x = 21,0 \pm 4,9$  en los ítems pares y  $21,5 \pm 5,0$  en los ítems impares); mientras que en la segunda aplicación resultó 0,92 ( $x$  de los ítems pares =  $20,0 \pm 4,8$  y  $20,7 \pm 4,6$  para los ítems impares). Asimismo, en el dibujo de una mujer, el coeficiente de Spearman-Brown fue de 0,92 en la primera aplicación ( $x = 20,1 \pm 5,0$  para los ítems pares y  $22,2 \pm 6,3$  para los impares); en la segunda aplicación la confiabilidad interna fue de 0,89 ( $x = 20,0 \pm 4,0$  para los ítems pares y  $20,4 \pm 4,1$  para los impares).

Estos datos indican que el instrumento está estructurado de tal manera que de 23 partes de la figura masculina, 15 están evaluadas en ambas mitades de la prueba, además, de la ropa, el perfil, la coordinación motora, la calidad de las rayas y la forma de la figura. En el dibujo de una mujer ocurre lo mismo: el cuello está evaluado en un ítem par y otro impar; la nariz, el pelo, los dedos de la mano y los pies están evaluados en cuatro ítems cada uno, de los cuales dos responden a ítems pares y dos a impares; en el caso de la ropa, son seis ítems divididos en tres para cada mitad. Como consecuencia, el coeficiente de confiabilidad interna, calculado con la fórmula Spearman-Brown, fue alto (ver Cuadro 3).

El resultado con la fórmula Kuder-Richardson para definir la confiabilidad interna del dibujo de un hombre fue de 0,81 y 0,80 ( $x = 42,2 \pm 9,6$  y  $40,6 \pm 9,3$ ) en la primera y segunda aplicación respectivamente; en el dibujo de una mujer fue de 0,84 y 0,71 ( $x = 42,2 \pm 10,1$  y  $41,2 \pm 7,6$ ). Se pudo notar que los resultados son más bajos que los de la Spearman-Brown debido a que las desviaciones estándar son más altas. Esto quiere decir que hubo algunos niños con puntuaciones altas y otros con puntuaciones bajas, de modo que el grupo de la muestra fue heterogéneo.



En el Cuadro 4 se aprecia que se debe usar un error estándar de cuatro puntos. Esto quiere decir que la puntuación real oscila entre cuatro puntos arriba o abajo del puntaje obtenido por el niño en cualquiera de los dibujos y en el promedio de los dos.

### Conclusiones

1. La Prueba de Goodenough-Harris (PGH) es una medición científicamente objetiva que pueda ser aplicada a la población escolar costarricense. Presenta una  $r = 0,87$ .
2. La PGH es un instrumento confiable (estable) científicamente que puede ser utilizada con la población escolar costarricense. Presenta una  $r = 0,87$ .
3. La confiabilidad interna de la PGH, determinada por la fórmula Spearman-Brown, asegura que esta prueba genera intelectual en nuestra población escolar; con esta se obtuvieron coeficientes de 0,91 y 0,92 en el dibujo de un hombre y 0,92 y 0,89 en el dibujo de una mujer.
4. Los resultados, al calcular el coeficiente de Kuder-Richardson, fueron 0,81 y 0,80 en el dibujo de un hombre y 0,84 y 0,71 en el dibujo de una mujer (en la primera y la segunda aplicación respectivamente).
5. Los resultados de la fórmula empleada para determinar el error estándar están dentro de los límites fijados, ya que éste fue de cuatro puntos.

### Bibliografía

- Barton, K. *Recent Data on the Culture Fair Scales*. Information Bulletin No. 16. Champaign, III: Institute for Personality & Ability Testing, 1973. See footnote 4, Chap. 8, for a review of cross-cultural research with the Draw-A-Person technique.
- Cattell, P. *Cattell Infant Intelligence Scale*. The Psychological Corporation, New York, 1960.
- Cattell, R. *Culture Fair Intelligence Test Institute for personality and Ability*. Testing Champaign, Illinois, 1963.
- Ceciliano-Rogers, H. *Objetividad y confiabilidad de la prueba Goodenough-Harris en Costa Rica*. Tesis de Licenciatura. Heredia: Universidad Nacional, 1988.
- Davidoff, L. *Introducción a la psicología*. Segunda edición. México: Mc Graw-Hill, 1984.
- Golomb, C. Representational development of the human figure: a look at the neglected variables of SES, IQ, sex and verbalization. *Journal of genetic psychology*, 1977, 131: 207-222.
- Harris, D. B. *Children's drawings as measures of intellectual maturity*. New York: Harcourt, Brace and World, 1963.
- Lewis, H. P. y Livson, N. Personality correlates of IQ discrepancy: Stanford-Binet and Goodenough-Harris. *Journal of genetic psychology*, 1977, 131: 237-242.
- Lindeman, R. *Tratado de medición educacional*. —Primera edición.— Buenos Aires: Editorial Paidós, 1971.
- McCarty, D. *Manual for the McCarthy Scales of Children's Abilities*. New York: Psychological Corp, 1972.
- Mehrens W.A. y Lehmann I.J. *Medición y evaluación en la educación y en la psicología*. —Primera edición— México: Compañía Editorial Continental, S.A., 1982.
- Méndez, Z. y Stover L. *Prueba de madurez para la lectura*. Costa Rica: UCR. Facultad de Educación, San José, 1980.
- Naglieri, J. A. y Maxwell, S. Inter-rater reliability and concurrent validity of the Goodenough-Harris and McCarthy draw-a-child scoring systems. *Perceptual and motor skills*, 53: 343-534.
- Popham, James W. *Modern educational measurement*. —Primera edición— Englewood Cliffs, N. J.: Prentice-Hall, 1981.
- Rodríguez Barquero, Roberto. *Adaptación y validación de la escala KTRS para uso en el diagnóstico temprano de problemas de aprendizaje de lectura en Costa Rica*. San José: CEMIE-MEP-OEA, Publicación No. 124, Agosto, 1985.

- Szekely, B. *Los tests: manual de técnicas de exploración psicológica*. Buenos Aires: Kapelusz, 1960.
- Terman, L. M., G. M. A. Merrill. *Stanford-Binet. Intelligence Scale: Manual for the third Revision, Form L-M* Boston: Houghton Mifflin, 1960.
- Wechsler, D. Wechsler. *Pre-School and Primary Scale of Intelligence*. New York, Psychological Corporation, 1949.
- Wechsler, D. *Manual for The Wechsler Intelligence Scale for Children Revised*. New York: Psychological Corporation, 1974.
- Weis, S. C. Culture fair intelligence test and draw-a-persons score from rural peruvian sample. *Journal of social psychology*, 1980, 11: 147-148.
- Weis, S. C. Goodenough-Harris drawing test scores from two eastern peruvian sample. *Journal of social psychology*, 1981, 114: 159-166
- Woodburn, S. S. *El diagnóstico temprano de problemas de aprendizaje: Test de la Escuela Meeting Street*. Heredia: EUNA, 1984.
- Woodburn, S. S. y Méndez, C. A. *Prueba de diagnóstico de imagen corporal*, sin publicar.

### ANEXO

CUADRO No.1:

Media y desviación estándar de las puntuaciones empleadas para determinar la objetividad de la PGH, por sexo del dibujo, promedio y aplicación (N = 49)

	I aplicación		II aplicación	
	$\bar{x}$	D.S.	x	D.S.
Dibujo de un hombre	43,3	10,2	40,5	10,3
Dibujo de una mujer	42,4	11,2	40,9	8,3
Promedio de los dos dibujos	42,8	10,2	40,9	8,7

CUADRO 2

Media y desviación estándar de las puntuaciones empleadas para determinar la confiabilidad de la PGH, por sexo del dibujo, promedio y aplicación (N = 34)

	I aplicación		II aplicación	
	$\bar{x}$	D.S.	$\bar{x}$	D.S.
Dibujo de un hombre	41,5	8,3	40,9	7,9
Dibujo de una mujer	42,3	8,0	41,7	6,5
Promedio de dos dibujos	41,8	7,8	41,6	6,8

CUADRO 3

Resultados de la prueba T de student  
entre los dibujos y entre aplicaciones

	n=34	n=49	n=83
<i>Entre dibujos</i>			
I aplicación	0,40	0,41	6,89
II aplicación	0,45	0,21	0,45
<i>Entre aplicaciones</i>			
Dibujo de un hombre	0,30	1,35	1,30
Dibujo de una mujer	0,34	0,75	0,73

CUADRO 5

ERROR ESTANDAR DE LA PGH, POR SEXO DEL DIBUJO,  
PROMEDIO Y APLICACION

	I aplicación	II aplicación
<i>Objetividad (n=49)</i>		
Dibujo de un hombre	3,7	3,7
Dibujo de una mujer	4,0	3,0
Promedio de los dibujos	3,1	3,7
<i>Confiabilidad (n=34)</i>		
Dibujo de un hombre	3,4	3,3
Dibujo de una mujer	3,3	2,7
Promedio de los dos dibujos	3,2	2,8

CUADRO 4

Objetividad y confiabilidad de la PGH,  
según sexo del dibujo y puntuación total

	Dibujo de un hombre	Dibujo de una mujer	Calificación Promedio
<i>Objetividad (n=49)</i>			
Correlación de Pearson	0,82	0,80	0,87
<i>Confiabilidad externa (n=34)</i>			
Correlación de Pearson	0,78	0,82	0,87
<i>Confiabilidad interna (n=83)</i>			
Correlación de Spearman-Brown (I aplicación)	0,91	0,92	--
Correlación de Spearman-Brown (II aplicación)	0,92	0,89	--
Fórmula de Kuder-Richardson (I aplicación)	0,81	0,84	--
Fórmula de Kuder-Richardson (II aplicación)	0,80	0,71	--