LA FIDELIDAD DE LAS PRUEBAS DE INSTRUCCION POR EL METODO DE MITADES

Hace algún tiempo (1) tuve que refundir a modo de resumen datos referidos a elaboración de pruebas instructivas. Hoy creo necesario ampliar las nociones elementales de entonces.

Ya no basta con decir: al estudiar la fidelidad, o exactitud con la que un test mide la aptitud, por el método de mitades se disminuye el valor y al rectificarlo se eleva la estimación.

Fidelidad, constancia, fiabilidad, confianza, equivalencia, estabilidad, consistencia..., son términos o que pretenden traducir la original «reliability» o que se hallan ligados desde alguna perspectiva. Todos ellos dan por sentado que el primer aspecto de la validez, medir lo que pretende medir, se ha cumplido. Pero cuando con cualquier unidad métrica intentamos una medida podremos medir perfecta o imperfectamente. La medida será perfecta si el instrumento utilizado también lo es y si la medición no adolece de error o discrepancia. En caso contrario será imperfecta.

Una prueba es fiel o de gran fidelidad cuando los resultados obtenidos representan la verdadera puntuación del examinando. Es constante o estable, porque al ser aplicada por segunda o tercera vez la relación entre las puntuaciones de los examinados (eliminado el factor aprendizaje), se mantiene estadísticamente similar. Es de confianza o gran fiabilidad porque dadas la exactitud y precisión de medida podemos estar seguros de que su uso producirá los resultados genéricamente esperados.

La experiencia ha demostrado que, hasta ahora, la prueba ideal de máxima fidelidad, fiabilidad o constancia no existe. Su inexistencia no supone una posición pesimista ya que la aproximación es cada vez mayor. En este acercamiento contribuyen todas las técnicas auxiliares que tienden a estudiar las causas de error y los medios de reducción de la discrepancia. Aunque el coeficien-

(1) Elaboración empírica y científica de las pruebas de instrucción. *Bordón*, núm. 21, 1951, págs. 241-245.

te 1 (máximo obtenible) esté lejos de nosotros, podemos constatar el sentido aproximativo de los hallazgos experimentales respecto de algunas pruebas. En este acercamiento es donde se halla el problema en torno a la fidelidad examinadora, con todos los matices diferenciales que hoy presenta.

Entre las técnicas que más facilitan el estudio de los errores encontrables en una prueba tenemos el análisis de varianza ideado por Fisher (1923) y perfeccionado posteriormente. Apoyado en la distribución normal o casi normal, nos aproxima en el estudio de la fiabilidad hasta el punto que una de las definiciones de fidelidad es: proporción de la varianza que es verdadera varianza (2). Se facilita la aproximación porque al estimar la varianza en distintos componentes se puede averiguar el influjo de cada uno. Con las técnicas oportunas se podrá determinar la significación de tales influjos.

El término varianza, no muy utilizado en tratados anticuados, significa el cuadrado de la desviación típica de la población. Como ésta suele ser desconocida, se estima de la muestra con estimadores insesgados.

Como es muy conocido la varianza total es partible en distintos componentes. Así cuando hay independencia entre los resultados de las diferentes partes, la fórmula de partición es:

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots\ldots + \sigma_{k-1}^2 + \sigma_k^2 = \sum_{i=1}^{k} \sigma_i^2$$

Si esta independencia no existiese la fórmula general es

$$\sigma^2 = \sum_{i=1}^{k} + 2 \sum_{i,j} \rho_{ij}\, \sigma_i\, \sigma_j \qquad \left[\begin{matrix} i, j, = 1, 2, \ldots.\, k-1, k \\ i < j \end{matrix}\right]$$

En las que σ_1^2, σ_2^2.... representan la cuantía de varianza correspondiente a la parte o aspecto 1, 2 ... y $\rho_{ij}\, \sigma_i\, \sigma_j$ la covarianza entre las partes i, j.

Quizá la comprensión fuese más clara si nos refiriésemos a puntuaciones. Así la puntuación X_{is} del sujeto *is* en la prueba pueda considerarse descomponible en

$$X_{is} = \overline{X}_1 + \overline{X}_2 + \overline{X}_3 + \ldots\ldots + X_z$$

(2) De ahora en adelante emplearemos indistintamente los términos fidelidad, fiabilidad o constancia.

$\overline{X}_1$ corresponde a la puntuación media de todos los escolares de su edad.

$\overline{X}_2$ corresponde a la diferencia entre la media de los alumnos de su mismo cociente intelectual y la media general.

$\overline{X}_3$ corresponde a la diferencia entre la media de los alumnos del mismo aprovechamiento escolar y la media de los alumnos de su mismo cociente intelectual.

Finalmente, en este ejemplo X_z representa la parte de puntuación correspondiente a un aspecto o peculiaridad imprevisible. Si se quiere podría afirmarse que corresponde a la «originalidad» del sujeto o podría descomponerse en originalidad y otros aspectos fácilmente resumibles con el concepto de error.

En correspondencia con lo anterior las varianzas σ^2_I, σ^2_1 serían las varianzas correspondientes a cada uno de los aspectos anteriores. El problema consistirá en determinar la partición y el influjo de cada parte respecto de las demás y respecto del todo.

El problema de la partición es muy sugeridor por ser múltiples los orígenes de las varianzas sistemática y fortuita. Ofrece algunas dificultades para la exacta distinción entre ambas y tiende a precisarse cuando hay numerosos estudios respecto de cada una. Pero nos aleja momentáneamente del tema.

Supuesto que σ^2_k sea la discrepancia o error inexplicable por las k — 1, partes sistemáticamente conocidas, y si llamásemos σ^2_∞ a la suma de las varianzas de todos los factores determinables con precisión tendríamos que

$$\sigma^2 = \sum_{i=1}^{i=k-1} \sigma^2_i + \sigma^2_k = \sigma^2_\infty + \sigma^2_k$$

Se comprende que cuanto menor sea σ^2_k (σ^2_e según la nomenclatura más comprensiva), es menor el influjo de lo impredecible y aumenta la exactitud de la medida. Es corriente denominar a σ^2_∞ varianza de las verdaderas puntuaciones y a σ^2_e varianza de los errores fortuitos.

El coeficiente $r_{nn} = r_{tt} = r_{II}$ (en otros términos correlación de un test consigo mismo, a través de las muestras experimentales), se obtiene así:

$$r_{nn} = \frac{\sigma^2_\infty}{\sigma^2} = 1 - \frac{\sigma^2_e}{\sigma^2}$$

Mas el concepto de fidelidad surgió mucho antes (1910) que el análisis de la varianza. Apareció unido al de correlación entre dos mitades de las puntuaciones de medidas de la misma cosa. El estudio de su error probable fué bastante posterior (1925). Debido a esto el concepto de fiabilidad queda unido históricamente al método de mitades.

El razonamiento lógico en que se apoya el método de mitades viene a ser este: En todo test ordenado la puntuación obtenida por los sujetos en los elementos impares (por ejemplo), debe mantenerse en relación constante con la puntuación obtenible en los elementos pares. Toda discrepancia es atribuible a factores incontrolados por la prueba experimentada.

La concordancia perfecta indica homogeneidad y exactitud, indica que dentro del mismo test no hay elementos susceptibles a las fluctuaciones fortuitas, pero seguras de los sujetos examinados. Su aparición nos obligaría a revisar el estudio.

El método de mitades con su mismo nombre indica se inicia mediante la partición de las puntuaciones de cada sujeto en dos grupos correspondientes al total de aciertos acumulados en cada mitad de los elementos previamente separados del test. La correlación entre ambos totales servirá de punto de apoyo para la determinación de la fidelidad. Se podría afirmar que constituye un caso particular de correlación entre partes de cualquier extensión o bien entre partes en número indeterminado.

Al hablar de método de mitades hemos de recordar que el número de particiones posibles depende exclusivamente del número de elementos de la prueba.

La fórmula combinatoria que nos da el número de partes es (si hacemos N = 2n y N/2 = n).

$$C_{2n,n} = \frac{V_{2n,n}}{P_n} = \frac{(2n)!}{n!\,n!}$$

dividido entre dos ya que son correlacionados.

Con un test de 100 elementos obtendríamos el siguiente astronómico resultado:

$$\text{Log } C_{2n,n} = \log (2n)! - [\log 2 + 2 \log n!] = 157{,}9700037 -$$
$$- (0{,}30103 - 2 \times 64{,}4830749) = 28{,}70282$$

El antilogaritmo tendría 29 cifras enteras. Entraría en el coto de lo quinquillones. Prácticamente en este caso el número de mitades es indefinido.

Ante la imposibilidad de elegir todas las muestras posibles y obtener la distribución de las intercorrelaciones para determinar la verdadera correlación entre mitades se suele recurrir a procedimientos prácticos. Criterios que por su sencillez convierten en una operación fácil la de totalizar la puntuación correspondiente a cada una de las mitades.

Entre estos tenemos:

a) Pares-impares o elementos alternos

b) Mitades completas

c) Grupos de elementos que totalicen las dos mitades.

d) Elementos de dificultad y discriminación equivalentes: Menos práctico.

Todos ellos estarán sometidos a crítica. De momento hacemos constar la inexistencia de alguna prueba de fidelidad que no pueda ser considerada como deficiente o ineficiente en ciertos casos. El criterio acumulativo de concordancia de todas las pruebas posibles: mitades, equivalencia, repetición, equivalencia racional, análisis de varianza, consistencia intrínseca, sensitividad... o de algunas de ellas, fortalecerá las conclusiones, aunque incremente el trabajo.

Derivación de la fórmula de predicción

Dijimos ya que la correlación entre las puntuaciones que totalizan cada mitad es el punto de partida de la fórmula de profecía de Spearman-Brown. Veámoslo. Si la puntuación X_i del sujeto *i* la consideramos descompuesta en:

$$Xi = (X_1 + X_2) + (X_3 + X_4)$$

donde X_1, X_2, X_3 y X_4 representan la puntuación en cada cuarta parte.

Tendremos para $x_i = X_i - \overline{X}, \ldots\ldots$

$$r_{(x_1+x_2)(x_3+x_2)} = \frac{\Sigma(x_1+x_2)(x_3+x_4)}{N\,\sigma_{(x_1+x_2)}\,\sigma_{(x_3+x_4)}} =$$

$$= \frac{\Sigma x_1 x_3 + \Sigma x_1 x_4 + \Sigma x_2 x_3 + \Sigma x_2 x_4}{N\left[\left(\sigma^2_{x_1}+\sigma^2_{x_2}+2\,r_{x_1 x_2}\,\sigma_{x_1}\,\sigma_{x_2}\right)\left(\sigma^2_{x_3}+\sigma^2_{x_4}+2\,r_{x_1 x_3}\,\sigma_{x_1}\,\sigma_{x_3}\right)\right]^{\frac{1}{2}}} =$$

(para simplificar: x_1 por 1, x_2 por 2,......

$$= \frac{r_{13}\,\sigma_1\,\sigma_3 + r_{14}\,\sigma_1\,\sigma_4 + r_{23}\,\sigma_2\,\sigma_3 + r_{24}\,\sigma_2\,\sigma_4}{\sqrt{\sigma_1^2+\sigma_2^2+2\,r_{12}\,\sigma_1\,\sigma_2}\ \sqrt{\sigma_3^2+\sigma_4^2+2\,r_{34}\,\sigma_3\,\sigma_4}}$$

si ahora simplificásemos mediante el artificio de admitir que

$$x_1 = x_2 \text{ y } x_3 = x_4$$

obtendríamos que

$$r_{13} = r_{23} = r_{14} = r_{24} \qquad \sigma_1^2 = \sigma_2^2 \text{ y } \sigma_3^2 = \sigma_4^2 \text{ y si } r_{12}\,\sigma_1\,\sigma_2 = r_{13}\,\sigma_2\,\sigma_4$$

todos los $r_{ij}\,\sigma_i\,\sigma_j$ serían iguales.

La fórmula anterior se transformaría en:

$$r_{(1+2)(3+4)} = \frac{4\,r_{13}\,\sigma_1\,\sigma_3}{\sigma_1^2+\sigma_3^2+2\,r_{13}\,\sigma_1\,\sigma_3}$$

Bajo la hipótesis de homogeneidad de varianzas en las medias partes $\sigma^2(1+2) = \sigma^2(3+4)$ que desde ahora denominaremos 1, 2, obtendríamos:

$$r_{(1)(2)} = \frac{4\,r_{12}\,\sigma_1^2}{2\,\sigma_1^2\,(1+r_{12})} = \frac{2\,r_{12}}{1+r_{12}}$$

Es decir que hemos alcanzado la fórmula en mitades de Spearman-Brown. La generalización, demostrable por la misma vía, cuando las partes son del mismo número de elementos, nos da:

$$r_{nn} = \frac{n\,r_{11}}{1+(n-1)\,r_{11}}$$

Y la sobregeneralización de la anterior cuando no es idéntico el número de elementos produce la siguiente expresión:

$$r_{ab} = \frac{a\,b\,r_{12}}{\sqrt{a+(a^2-a)\,r_{12}}\ \sqrt{b+(b^2-b)\,r_{12}}}$$

Fórmula general de menor uso práctico, pero que por su generalidad es transformable en las anteriores o modificable en otras de diferente utilidad.

La determinación de la desviación típica de la fórmula r_{nn} recién presentada se lleva a cabo mediante la aplicación del siguiente hallazgo:

$$\sigma r_{nn} = \frac{n\left(1 - r^2_{II}\right)}{\sqrt{N}\,[1 + (n-1)\,r_{II}]^2} = \frac{n\,\sigma\,r_{II}}{[1 + (n-1)\,r_{II}]^2}$$

Modernamente ha avanzado la ciencia estadística al utilizar nuevos procedimientos. Así la aplicación del criterio de máxima verosimilitud ha producido nuevas fórmulas conforme las hipótesis admitidas. La correlación entre mitades que exige las hipótesis de $\sigma_1 = \sigma_2$ " $\mu = r_1$ ha producido una nueva estimación r' que a veces altera en más de cinco centésimas el valor anterior. No debe creerse que esto es insignificante ya que en muchos casos podrá suponer la aceptación o eliminación de nuevas hipótesis. Desde el punto de vista científico se considera como más rigurosa la determinación por el método de máxima verosimilitud.

La fórmula de aplicación obtenida con el nuevo procedimien to es:

$$r' = \frac{2\,\Sigma\,X\,Y - \dfrac{(\Sigma\,X + \Sigma\,Y)^2}{2\,N}}{\Sigma\,X^2 + \Sigma\,Y^2 - \dfrac{(\Sigma X + \Sigma Y)^2}{2\,N\,Y}}$$

Bajo las consideraciones anteriores y antes de comentar la suficiencia o insuficiencia del procedimiento junto a sus vías de aplicabilidad eficiente, presentamos en la Tabla 1.ª un supuesto de estudio de la fidelidad de un test de cuarenta elementos por el método de mitades.

Para mayor claridad, representaremos el acierto en un elemento por medio de una cruz y el error por un guión.

Entre las posibles mitades obtenibles hemos trabajado con las tres siguientes (pares-impares) (primera y cuarta decenas con segunda y tercera) y (primera y tercera decenas con segunda y cuarta). De este modo se puede comprender mejor el conjunto de posibilidades que favorezcan o perjudiquen cierta agrupación. También hemos aplicado como contraste el criterio abreviado de Rulon.

TABLA I.- Análisis de respuestas pa

Sujeto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	2
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	-	+	+	+	+	+	-	+	-
4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-	-	-	-	+	+	+	-	-
5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-	+	-	+	+	-	+	-	+	-
6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	-	+	+	-	-	-	+	+	-	-
7	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	+	+	+	-	+	-	+	-
8	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	-	-	+	-	+	+	+	-	-
9	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+	+	-	-	-	-	+	+	-	-
10	+	+	-	+	+	+	+	+	-	+	+	+	-	+	+	+	+	+	-	+	+	+	-	-	+	+	-	+	-
11	+	+	+	+	+	-	+	+	+	+	+	+	-	+	+	-	+	+	-	+	-	+	+	-	-	-	+	-	-
12	+	+	+	+	+	+	-	+	+	+	+	+	-	+	+	-	+	+	-	+	-	+	+	+	+	-	-	-	-
13	+	+	+	+	+	+	+	-	+	+	+	-	-	+	-	+	+	+	+	+	-	+	+	+	-	-	-	-	-
14	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-
15	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-
16	+	+	+	-	+	+	+	+	-	+	+	+	-	+	+	-	+	+	+	-	+	+	-	+	-	-	+	-	-
17	+	+	+	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+	+	-	+	+	-	-	+	-	+	-	-
18	+	+	+	+	-	+	+	+	+	-	+	+	+	-	+	-	+	+	+	-	+	+	-	+	-	-	+	-	-
19	+	+	+	+	+	+	-	+	+	+	+	-	+	+	-	+	+	+	+	-	-	+	-	-	-	-	+	-	-
20	+	+	+	+	+	-	+	+	+	+	-	+	+	+	-	+	+	-	+	-	+	+	-	-	-	-	+	-	-
21	+	+	-	+	+	+	+	+	-	+	+	-	+	-	+	+	+	+	+	+	-	+	-	-	+	-	+	-	-
22	+	+	+	+	+	+	-	+	+	-	+	-	+	-	+	-	+	+	-	+	+	-	+	-	-	-	-	+	-
23	+	+	+	+	+	+	+	-	+	+	-	+	-	+	+	+	-	+	-	+	-	-	+	+	+	-	-	-	-
24	+	+	+	+	-	+	+	+	+	-	+	+	-	+	+	+	-	+	+	+	-	-	-	+	-	+	-	-	-
25	+	+	+	+	+	+	+	-	+	+	-	+	+	+	-	-	+	-	-	-	+	+	+	+	+	-	-	-	-
26	+	+	+	+	+	-	+	+	+	+	-	+	+	+	-	+	-	-	+	-	+	-	+	-	-	+	-	+	-
27	+	+	+	-	+	+	+	+	-	+	+	+	+	-	+	-	-	-	-	-	+	+	+	+	+	-	-	+	-
28	+	+	+	+	+	+	-	+	+	+	-	+	+	+	-	+	-	-	+	-	+	-	-	-	-	+	-	+	-
29	+	+	+	+	+	+	+	-	+	+	-	+	+	+	+	-	+	-	+	-	-	-	+	-	-	-	-	+	-
30	+	+	+	+	+	-	+	+	+	-	+	+	+	-	+	+	-	-	+	+	-	-	+	-	+	-	-	-	-
31	+	+	+	+	-	+	+	+	+	+	-	+	-	+	+	-	-	+	-	+	-	-	-	+	-	-	-	-	-
32	+	+	+	+	+	+	+	-	+	-	+	+	+	-	+	+	-	-	-	+	-	-	-	-	+	-	-	-	-
33	+	+	+	+	+	+	+	+	-	+	-	+	+	-	+	-	-	-	-	+	+	-	-	-	-	+	-	-	-
34	+	+	+	+	+	-	+	+	+	-	+	-	+	+	-	+	-	-	-	+	-	-	-	+	-	-	-	-	+
35	+	+	+	+	+	+	-	+	-	+	+	-	+	+	-	+	-	-	-	+	-	-	-	-	-	-	-	+	-
36	+	+	+	-	+	+	+	+	-	+	+	-	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-
37	+	-	+	+	-	+	+	+	+	-	+	-	+	-	-	+	-	-	+	-	+	-	-	-	+	-	-	-	-
	37	36	35	34	33	32	32	31	30	29	28	27	27	26	25	24	23	22	21	20	19	18	17	16	16	15	14	13	1

studio de la fidelidad de las pruebas

32	33	34	35	36	37	38	39	40	Cuota del sujeto	Pares-Impares		1.° y 3.°	2.° y 4.°	1.° y 4.°	2.° y 3.°	Sujeto
+	+	+	−	+	−	−	−	+	36	17	19	20	16	16	20	1
+	+	−	+	−	−	−	+	−	35	19	16	20	15	15	20	2
−	−	+	+	−	+	−	+	−	30	17	13	17	13	15	15	3
+	+	−	+	−	+	+	−	−	29	16	13	15	14	15	14	4
−	−	−	−	+	+	+	−	−	28	14	14	16	12	14	14	5
−	+	+	+	−	−	+	−	−	27	13	14	15	12	14	13	6
+	−	−	−	+	−	−	−	−	26	12	14	16	10	12	14	7
+	−	−	−	−	+	−	−	−	25	13	12	15	10	12	13	8
−	−	−	+	−	−	−	−	−	24	14	10	14	10	12	12	9
−	−	−	−	−	−	−	−	−	23	10	13	14	9	9	14	10
−	−	+	+	−	−	−	−	−	22	12	10	12	10	12	10	11
−	−	−	−	−	−	−	−	−	21	9	12	14	7	9	12	12
−	−	−	−	+	−	−	−	−	21	9	12	13	8	10	11	13
−	−	−	−	−	−	−	−	−	20	10	10	10	10	10	10	14
−	−	−	−	−	−	−	−	−	20	10	10	20	0	10	10	15
+	−	−	−	−	−	−	−	−	20	10	10	12	8	9	11	16
+	−	−	−	−	−	−	−	−	20	13	7	12	8	9	11	17
−	−	−	−	−	−	−	−	−	19	11	8	12	7	8	11	18
−	−	+	−	−	−	−	−	−	19	9	10	11	8	10	9	19
−	+	−	−	−	−	−	−	−	19	11	8	12	7	10	9	20
−	−	−	−	−	−	−	−	−	19	10	9	11	8	8	11	21
−	+	−	−	−	−	−	−	−	19	12	7	12	7	9	10	22
+	−	−	−	−	−	−	−	−	19	8	11	12	7	10	9	23
+	−	−	−	−	−	−	−	−	19	7	12	10	9	9	10	24
−	−	−	−	−	−	−	−	−	19	11	8	14	5	10	9	25
−	−	−	−	−	−	−	−	−	18	9	9	13	5	9	9	26
−	−	−	−	−	−	−	−	−	18	10	8	14	4	8	10	27
−	+	−	−	−	−	−	−	−	18	8	10	12	6	10	8	28
−	−	+	−	−	−	−	−	−	18	10	8	11	7	10	8	29
−	+	−	−	−	−	−	−	−	18	12	6	10	8	9	9	30
−	−	−	−	−	−	−	−	−	17	6	11	12	5	9	8	31
−	−	−	−	−	−	−	−	−	16	9	7	10	6	8	8	32
−	−	+	−	−	−	−	−	−	16	7	9	11	5	10	6	33
−	−	−	−	−	−	−	−	−	16	9	7	10	6	9	7	34
−	−	−	−	−	−	−	−	−	15	5	10	10	5	8	7	35
−	−	−	−	+	−	−	−	−	15	7	8	9	6	9	6	36
−	−	−	−	−	−	−	−	−	14	10	4	9	5	8	6	37
9	8	7	6	5	4	3	2	1	778	399	379	480	298	384	394	

Conforme dijimos el método de mitades exige la existencia de ciertas condiciones o hipótesis. Luego hemos de comprobar en las muestras obtenidas si dichas hipótesis son ciertas.

En primer lugar estudiamos la homogeneidad de las varianzas de las seis partes. En la Tabla II aparecen resumidos los cálculos para dicha obtención, mediante el criterio L_1.

Tabla II. - Determinación de la homogeneidad de varianzas

n_m	$\log n_m$	$n_m \log n_m$	θ'_m	$\log \theta'_m$	$n_m \log \theta'_m$
37	1'56820	58'02340	352'2703	2'54688	94'23456
37	1'56820	58'02340	322'8108	2'50796	92'79452
37	1'56820	58'02340	308'9730	2'48868	92'08116
37	1'56820	58'02340	383'8919	2'58329	95'58173
37	1'56820	58'02340	188'7027	2'27580	84'20460
37	1'56820	58'02340	392'4325	2'59333	95'95321
222		348'14040	1949'0812		554'84978
$\log L_1 = 2'34635 - 1'56820 + 2'49932 - 3'28979 = -0'01232 = \bar{1}'98768$					

Los cálculos y la comparación en las Tablas correspondientes nos permiten aceptar la hipótesis de que las varianzas de las partes son iguales. La diferencia apreciada a primera vista se debe sólo a las fluctuaciones muestrales.

La segunda de las hipótesis la verificamos mediante el análisis de varianza resumido en la Tabla III. El estudio del mismo nos ofrece una primera muestra de los contrastes estadísticos. Aunque hubiésemos aceptado el supuesto de identidad formal (3), lo cierto es que las medias obtenidas destruyen la hipótesis de homogeneidad por el influjo de las agrupaciones presentadas.

(3) FERNÁNDEZ HUERTA, J.: El criterio intrínseco en la hipótesis de normalidad dentro de la experimentación didáctico-pedagógica. *Revista Española de Pedagogía*, núm. 44, octubre-diciembre 1953, págs. 517-527.

— Puntuaciones tipificadas en la composición de notas. *Revista Española de Pedagogía*, núm. 46, abril-junio 1954, págs. 153-156.

Tabla III. - Homogeneidad de medias

	Grados de libertad	Suma de cuadrados	Cuadrado medio	F	Hipótesis 0,1 %
Entre medias.........	5	454	90'8	10'007	Rechazada
Dentro de las partes...	216	1949	9'02		
TOTAL.........	221	2403			

Obtenidas las correlaciones por el procedimiento habitual

$$r = \frac{\Sigma x y}{N \sigma_x \sigma_y}$$

y por el método de máxima verosimilitud los resultados pueden apreciarse en la tabla siguiente:

!!	A	B	C	Medias
r	0,47	0,43	0,75	0,55
r'	0,47	—0,12	0,72	0,36
z'	0,51	0,46	0,97	0,65

La anomalia de *r* y *r'* en B puede atribuirse a la falta del cumplimiento de las condiciones para máxima verosimilitud en tal modo de agrupar.

Para determinar si la obtención de las tres muestras puede o no producir correlaciones que sean significativamente distintas, realizaremos el método más breve que, aceptado, nos reduce cálculos posteriores. Compararemos las correlaciones de B y C por ser las de mayor diferencia.

En el caso de las *r*, el valor de la RC es significativo el 1 por 1.000. Igual acontece con otras comparaciones de esta Tabla.

La prueba de significación obtenida para las *r*, excluído el — 0,12, nos da una probabilidad superior al 1 por 100. Luego podemos tener certeza muy significativa.

Bajo este supuesto de existencia real de las *r* se pueden aplicar las anteriores fórmulas.

De este modo la fiabilidad del test de 40 elementos presentado y determinada mediante la experimentación aquí resumida sería, para caso pares-impares:

$$r_{nn} = \frac{2 \times 0{,}47}{1 + 0{,}47} = 0{,}64$$

A partir de este 0,64 podremos arbitrariamente bien determinar la fidelidad de un test de instrucción de las mismas características, pero de mayor número de elementos, bien el número de elementos necesario para alcanzar una fiabilidad determinada.

De un modo práctico, pero sujeto a error, lo podríamos realizar con el gráfico de la figura primera. Bastaría averiguar gráficamente la ordenada correspondiente a cierta abscisa que sería igual a la distancia entre la curva correspondiente al r_{nn} encontrado. En el otro caso se hallaría la abscisa dada la ordenada.

Si quisiéramos predecir la fidelidad de un test de 100 elementos, operaríamos así: $n = 100 : 40 = 2'5$. Luego el nuevo tests es 2'5 veces mayor que el original. Mediante el gráfico encontraríamos rápidamente que la correlación predicha se situaría alrededor de 0,83.

Mediante la fórmula hubiésemos obtenido con precisión el valor de

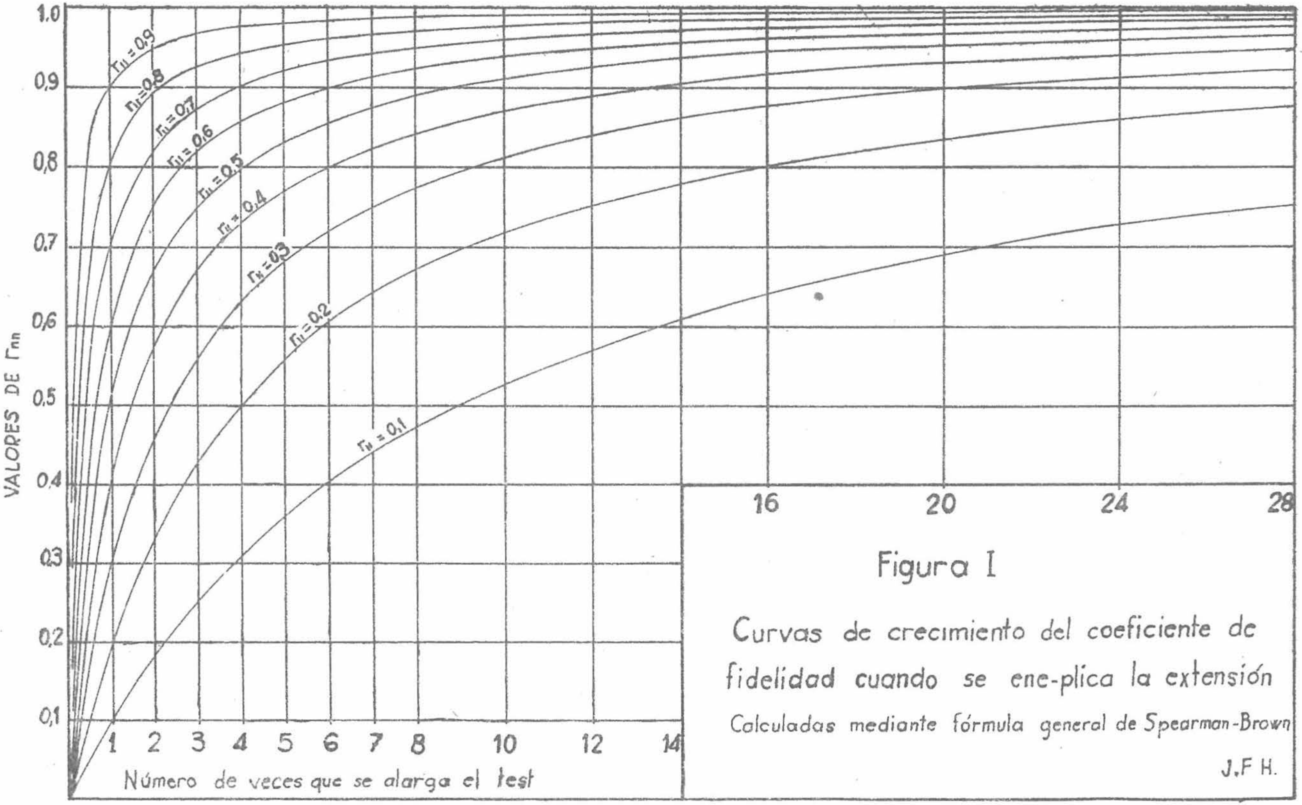
$$r = 0{,}816 \backsimeq 0{,}82$$

Para determinar el número de elementos necesarios para cierta correlación dada aplicaríamos la fórmula derivada

$$n = \frac{r_{nn}\,(1 - r_{II})}{r_{II}\,(1 - r_{nn})} \qquad \sigma_n = \frac{n\;(1 + r_{II})}{r_{II}\sqrt{N}}$$

Si buscásemos la correlación de 0'90, mínimo exigible en estos casos, alcanzaríamos un $n \backsimeq 5$. Luego el número de elementos necesarios sería el de 200. Pronto se advierte la excesiva cantidad de preguntas a las que deberíamos someter a los escolares. El gráfico nos hubiera dado un número aproximado a 5, en todo muy similar al recién hallado.

Por vía de comparación con la definición de fidelidad como relación entre la varianza de las puntuaciones verdaderas y la va-


1.0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
VALORES DE r_{nn}
r_{11} = 0,9
r_{11} = 0,8
r_{11} = 0,7
r_{11} = 0,6
r_{11} = 0,5
r_{11} = 0,4
r_{11} = 0,3
r_{11} = 0,2
r_{11} = 0,1
1 2 3 4 5 6 7 8 10 12 14
16 20 24 28
Número de veces que se alarga el test
Figura I
Curvas de crecimiento del coeficiente de fidelidad cuando se ene-plica la extensión
Calculadas mediante fórmula general de Spearman-Brown
J.F H.

rianza total, aplicaremos el procedimiento a partir de los datos tabulados.

Aplicada como ejemplo a una sola partición: A, obtendremos un $r_{rr} = 0{,}468$.

Tal resultado se distancia bastante del 0,64 hallado, debido a razones que sistematizaremos a continuación.

Crítica del método de mitades

Una de las primeras ventajas que ofrece el método de mitades es la de reducir la experimentación en la elaboración de los tests instructivos a una sola serie de ensayos. Preparación, ensayos, corrección..., quedan reducidas a la mitad.

Otra de sus conveniencias radica en la posibilidad de ser aplicado con éxito a procedimientos de evaluación: juicios. Lo que hicimos en nuestra Escala de escritura (4). La relación puede exigir su mantenimiento o su partición.

La fórmula de mitades provee una correlación más alta que las formas paralelas o equivalente por suprimir la varianza de error propia del examen en diferentes situaciones y estado de ánimo. No toda persona reacciona ante un test del mismo modo en ocasiones diversas. Esta diversidad, prácticamente inapreciable, influye como origen de error fortuito y disminuye el coeficiente de fiabilidad. Se ha comprobado que la correlación entre las pruebas mentales Stanford, disminuía al aumentar la distancia entre los experimentos desde o'889 con el intervalo de un mes hasta o'698 con intervalo de cinco años. Otros experimentos han permitido defender la teoría de Spearman que acepta las variaciones en períodos cortos pero no en largos.

Precisamente esta variación en períodos cortos es una de las principales objeciones, totalmente insalvable, contra el método de mitades. Toda prueba aplicada durante una sola sesión pierde las variaciones diurnas. Y si se utiliza el procedimiento de pares-impares o elementos alternos, el influjo de esta variación queda mi-

(4) FERNÁNDEZ HUERTA, J.: *Escritura (Didáctica y escala gráfica)*. C. S. I. C., Madrid, 1950.

nimizado. Conforme hemos repetido al disminuir el error de fortuidad se incrementa la varianza de puntuaciones verdaderas y el test ofrece una fiabilidad superior a la que debiera mostrar. La estimación en este caso pecaría de excesiva, pero no se puede por ahora contrastar cuál es la cuantía de ese exceso.

Otra censura se ofrece al método de pares-impares: la correlación existente entre los resultados de elementos sucesivos a causa de su su semejanza estructural o contigüidad temporal. Esta correlación que nos exige considerar la fórmula más general para casos de no independencia, se traduce en nueva hiperestimación, ya que el aumento de varianza sistemática originado o disminución de la fortuita se transforma en un cociente mayor. Si nos fijamos en nuestro supuesto advertimos que la correlación máxima es precisamente la correspondiente a grupos alternos, y aunque científicamente sean todas las correlaciones de igual valía, mucho hubiese variado el poblema de realizar la estimación sólo a base de una de las particiones. Esta dificultad de influjo de la partición, junto a la posición accidental de los elementos en la prueba, es general para todo el método.

Por esto aconsejamos un procedimiento semejante al expuesto: combinación de elementos alternos y grupos alternos. Es francamente superior al que pudiéramos desechar de correlacionar la primera y segunda mitad. La dificultad teórica de este último procedimiento se fundamenta en:

El método de mitades no puede ser aplicado a tests de velocidad ya que los supuestos de estos tests son preguntas de dificultad asequible y similar junto a tiempo menor al que la persona más rápida podría utilizar para su realización. Solo es valioso cuando la unidad es el tiempo. En este caso el tiempo total se escinde en dos mitades separables o por detención o por trazo del alumno a orden intermedia.

El método de mitades es aplicable a los tests de profundidad o aprovechamiento, es decir, a los tests, en lo que el tiempo es teóricamente ilimitado, pero las preguntas deben ir ordenadas en orden creciente de dificultad. Luego los sujetos normalmente contestarán a más preguntas de la primera que de la segunda mitad. No tendría nada de particular, dado ese caso, que admitamos las siguientes hipótesis comprobables experimentalmente:

Los requisitos de homogeneidad exigibles: media y varianza no se cumplirán en la mayoría de los casos. Luego todo el proceso posterior quedaría rechazado.

Convendría que intentemos recordar todas las hipótesis que hemos comprobado o admitido en nuestro ejemplo:

a) Igualdad de la homogeneidad (varianzas y medias).

b) Igualdad de estructura formal.

c) Igualdad de las intercorrelaciones entre las partes divididas.

d) Todos los elementos son medidas de un solo factor (caso contrario produce hipoestimaciones).

Si se predice la fidelidad para nueva forma de tests alargado en varios elementos hay que contar con nuevas hipótesis:

e) Los elementos incluídos no reducen las caracteristicas propias del tests: contenido, forma, dificultad, consistencia interna, discriminación, estabilidad...

f) El alargamiento de la prueba no produce nuevas condiciones psíquicas que aumenten la varianza del error. No se debe olvidar que la presentación de pruebas de excesiva extensión requiere un tiempo que puede sobrepasar el óptimo para cada tests y reducir la fiabilidad.

José Fernández Huerta
Profesor de Didáctica de la Universidad
de Madrid

SUMMARY

One of the existing differences between the technical construction of achievement tests and the scientifical construction of the same is grounded on the determination of the reliability, a condition required for the last ones: A test is not scientifically accurate while its reliability is not known.

For this reason the author of this article points out the conceptual interpretation of that word both when it is translated into Spanish and in its own language. But having on account the extraordinary extent of the question he only adopts one of the most employed procedures: that of the split-halves.

He shows the corresponding theoretical aspects and presents a practical case which easily permits to follow the procedure. For the first time in the Spanish educational research he introduces the analysis of variance and the methods of maximum likelihood. He applies these concepts to the procedures of study of the reliability through the method of split-halves.