METODOS DE CONSISTENCIA Y EQUIVALENCIA EN LA DETERMINACION DE LA FIDELIDAD DE LAS PRUEBAS INSTRUCTIVAS

En un artículo muy reciente (1) estudiábamos la fidelidad de las pruebas instructivas por el método más antiguo y más utilizado: el de mitades.

Pero conforme indicábamos, a las ventajas propias de tal método se unian ciertos inconvenientes salvables cuando completamos el estudio de mitades por varias agrupaciones. Se podría asegurar que el promedio obtenido al realizar varias agrupaciones sería una de las más claras manifestaciones del coeficiente de fidelidad bajo el supuesto de una sola prueba y en una sola sesión.

Ahora bien, al procedimiento entonces bosquejado cabe añadir otra serie de métodos que pretenden resolver el problema de la fiabilidad o fidedignidad. La presentación de nuevos métodos no puede considerarse como algo superfluo, sino como una necesidad real. Cada uno de los sistemas empleados para averiguar la fidelidad de los «tests» está sometido a ciertas hipótesis y restricciones. Todos buscan la confianza en las pruebas realizadas y son realmente necesarios antes de admitir un «test» instructivo o psicológico como digno de fe.

Por esta razón y para completar el marco de las consideraciones en torno a la fidelidad, presentaremos ahora un conjunto de métodos de sumo interés. Definida la fidelidad como proporción de la varianza que es verdadera varianza, iniciaremos la consideración de un nuevo y sencillo método que ponga en juego las ideas del análisis de varianza y que nos ligue con el estudio de la fidedignidad por el método de mitades. Este *método* es el *de Rulon*. Se apoya dicho autor en el estudio de la varianza de las diferencias entre las puntuaciones logradas por cada sujeto en las dos mitades del

(1) «La fidelidad de las pruebas de instrucción por el método de mitades». *Revista Española de Pedagogía*, núm. 47, julio-septiembre, 1954, págs. 273-288.

test. Esta varianza se convierte en equivalente a la varianza del error y permite la obtención directa del coeficiente de fidelidad.

En sencillo esquema y sin explicar los supuestos que permiten el establecimiento de las igualdades, alcanzaríamos el siguiente proceso:

$$\sigma^2(x_1 - x_2) = \sigma^2_{x1} + \sigma^2_{x2} - 2\, r_{12}\, \sigma_{x1}\, \sigma_{x2} = \left(\sigma^2_{\infty 1} + \sigma^2_{e1}\right) + \left(\sigma^2_{\infty 2} + \sigma^2_{e2}\right) -$$

$$- 2\, r_{12}\, \sigma_{x1}\, \sigma_{x2} = \left(\sigma^2_{\infty 1} - r_{12}\, \sigma_{x1}\, \sigma_{x2}\right) + \left(\sigma^2_{\infty 2} - r_{12}\, \sigma_{x1}\, \sigma_{x2}\right) + \sigma^2_{e1} + \sigma^2_{e2} =$$

$$= 2\, \sigma^2_{e1} = \sigma^2_{et}$$

Conclusión que nos facilita la obtención de la fidelidad del siguiente modo:

$$r_{tt} = 1 - \frac{\sigma^2_{et}}{\sigma^2_t} = 1 - \frac{\sigma^2(x_1 - x_2)}{\sigma^2_t}$$

Del proceso anterior se infiere que si bien hemos operado con puntuaciones distribuídas en mitades las conclusiones valen para todo el test como unidad. La varianza de las diferencias entre las mitades equivale precisamente a la varianza del error, luego no necesitamos emplear fórmulas como la de Spearman-Brown.

Por no separarnos del ejemplo que ofrecimos al estudiar la fidelidad por método de mitades, presentaremos la misma tabla inicial. Tabla que aconsejamos para quienes hayan de estudiar la confianza de los instrumentos de medila. (Véase la tabla en las páginas 416-417.)

De acuerdo con los datos de dicha tabla, referidos solamente a la primera subdivisión, el valor hallado para r_{tt} es de 0,649.

Si recordamos la fidelidad por el método de mitades predicha mediante la fórmula de Spearman-Brown, para la mitad pares-impares que ahora hemos reconsiderado, advertimos la casi coincidencia de los resultados. Por esta causa el método de Rulon supone una ligera abreviación respecto del método general de mitades.

Dado que en el estudio de las formas equivalentes de un test se admite como condición para dicha equivalencia la identidad de las verdaderas varianzas, también puede aplicarse el método de Rulón para estimar la fidelidad con el método de pruebas equivalentes. Ahora bien, de acuerdo con las fórmulas presentadas, la varianza del error que admitiríamos como discriminativa, correspondería al conjunto de la prueba formada por las dos formas. La varianza total también habría de ser la de todas las puntuaciones conseguidas como suma del par de puntuaciones de cada sujeto. Por este motivo, lo que nosotros conseguiríamos sería la fidelidad de un test de doble número de preguntas. Para reducir la fidelidad a su valor ordinario, aplicaríamos la fórmula de predicción de Spearman-Brown, que en este caso sería:

$$r_{tt} = \frac{r_{12}}{2 - r_{12}}$$

De este modo no incrementaríamos excesivamente el valor de la fidelidad. La fidelidad del test sería inferior a la fidelidad del conjunto de los dos tests que habríamos obtenido de aplicar en toda su integridad el procedimiento de Rulón a las formas equivalentes.

Por su conexión y semejanzas con el método de Rulón, nos referiremos ahora a otro de los métodos más utilizados, cuando nos centramos primordialmente en las formas equivalentes.

Mentaremos en primer lugar el *método de sensitividad de Jackson*. En este método se aplican las técnicas del análisis de varianza para resolver varios problemas: 1. El efecto de la práctica sobre la equivalencia; 2. Su valor discriminativo entre los escolares, y 3. La sensitividad, es decir, la relación entre la desviación típica de las verdaderas puntuaciones y la de los errores de medida.

El procedimiento en sí no es complicado. Pero para alcanzar todos los procesos convenientes es necesario utilizar las siguientes fórmulas:

$$\gamma = \frac{\sigma_{\infty}}{\sigma_e} \qquad r_{tt} = \frac{\gamma^2}{1+\gamma^2} = \frac{\sigma^2_{\infty}}{\sigma^2_t}$$

No se establecen los procedimientos de varianza de la sensitividad ya que se prefiere la determinación de los límites fiduciales. Estos límites fiduciales vienen dados por las fórmulas

$$\underline{\gamma} = \sqrt{\frac{F}{2\,F_{0,01}} - \frac{1}{2}} \qquad \overline{\gamma} = \sqrt{\frac{F\,F_{0,01} - 1}{2}}$$

TABLA I.- Análisis de respuestas para

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	29
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 35 33 32 30 28 27 25 23 21 19 17 16 14 12

36 34 32 31 29 27 26 24 22 20 18 16 15 13

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	

Si a nuestro ejemplo le aplicamos el método de sensitividad de Jackson, obtendremos los siguientes valores presentados en la tabla de análisis de varianza:

TABLA II.—Análisis de varianzas.—Método de sensitividad de Jackson

ORIGEN DE VARIANZA	Grados de libertad	Suma de cuadrados	Cuadrado medio	F 1 %	HIPOTESIS
Efecto de la práctica...	1	5,41	5,41	1,11	Aceptada
Entre individuos........	36	449,48	138,87	2,84	Rechazada
Error de medida........	36	175,59	4,88	—	
TOTAL..........	73	680,48			

De este primer estudio podremos inferir que el valor de la práctica respecto de las dos formas del test, en nuestro caso dos medias formas, no influye como fuente de error sistemático, aunque actúe como origen de varianza fortuita. También nos permite asegurar que este tests produce diferencias significativas entre los sujetos. Es decir su valor discriminativo queda probado al estudiar la homogeneidad de la prueba.

La sensitividad del test γ es igual a 0,96 y sus límites fiduciales al 1 por 100 oscilan entre 0,35 y 1,63, lo que podríamos expresarlo así:

$$0{,}35 \leq \gamma \leq 1{,}63$$

Del mismo modo obtendríamos los valores medio y límites para la fidelidad de los test. Los valores que ahora conseguimos son:

r_{tt} **oscila entre** $0{,}12 \leq 0{,}48 \leq 0{,}73$

Quien se fije un poco rápidamente podrá encontrar un grave contraste entre este valor de la fidelidad y el determinado mediante los procedimientos Spearman-Brown y Rulon. Pero no debe llamarse a engaño. No podemos olvidar que ahora hemos operado bajo el supuesto de fidelidad en formas paralelas y en nuestro ejemplo hemos operado con mitades de tests. Luego nosotros hemos hallado la fidelidad del test mitad en extensión al propuesto. En este caso para averiguar la fidelidad del test total necesitamos emplear la fórmula predictiva de Spearman-Brown.

El valor obtenido al aplicar la fórmula de predicción nos produce un coeficiente de $r_{tt} = 0{,}649$.

Valor idéntico al obtenido mediante la fórmula de Rulon.

Podría preguntarse la razón que nos indica la conveniencia de utilizar este método de Jackson sobre el de Rulon, mucho más sencillo e igual de eficiente.

Las razones son varias. En primer lugar, nos permite determinar el efecto de la práctica respecto de las dos formas, y, en segundo lugar, nos proporciona seguridad de que el test discrimina entre los sujetos.

Más aún se precisa en el análisis de los componentes básicos en un test cuando con los mismos criterios de análisis de varianza se intenta determinar por una parte el influjo de las mismas preguntas en la fidelidad global y por otra el influjo de los sujetos. Con esta intención ideó *Hoyt su método de análisis de varianza* en el estudio de la fidedignidad.

Parte de la consideración general de varianza verdadera y varianza del error. Se desglosa la varianza verdadera en partes integrantes: La varianza atribuible a las preguntas, independiente del número de las mismas, y la varianza atribuible a los sujetos, independiente del número de los mismos. Conocida la varianza total se despeja la varianza del error y se procede a determinar la fidelidad por el método que tantas veces hemos recordado. Es decir, la fidelidad se obtendrá por la relación entre la verdadera varianza y la varianza de las puntuaciones.

Para la determinación de la tabla resumen III, en la que podremos estudiar con más facilidad el análisis de varianza, se deben emplear las fórmulas siguientes:

Suma de cuadrados entre individuos:

$$\frac{\sum\limits_{1}^{N} t_i^2}{n} - \frac{\left(\sum\limits_{1}^{N} t_i\right)^2}{n\,N}$$

Suma de cuadrados entre preguntas:

$$\frac{\sum_{1}^{n} p_i^2}{N} - \frac{\left(\sum_{1}^{n} p_i\right)^2}{n\,N}$$

Suma total de cuadrados:

$$\frac{\left(\sum_{1}^{N} t_i\right)\left(n\,N - \sum_{1}^{N} t_i\right)}{n\,N}$$

n = N.º de preguntas.
N = N.º de sujetos.
p_i = Total de aciertos en pregunta *i*.
t_i = Total de aciertos en sujeto *i*.

TABLA III.—Análisis de varianza.—Método de Hoyt.

ORIGEN DE VARIANZA	Grados de libertad	Suma de cuadrados	Cuadrado medio	$F_{1\%}$	HIPOTESIS
Entre examinados	36	24,9743	0,6937	4,35	Rechazada
Entre elementos	39	120,3217	3,0852	19,36	Rechazada
Residual o de error......	1.404	223,7283	0,1594		
TOTAL	1.479	369,0243			

Mediante este análisis podemos averiguar tres objetivos: *a*) El valor discriminativo general del test respecto de los sujetos examinados; *b*) El diferente grado de dificultad de las preguntas que integran el test, y *c*) La fidelidad del mismo test. El primero de los aspectos era resuelto con el método de Jackson, pero no el segundo. La importancia del segundo es básica cuando se intenta determinar si podemos considerar el test como de profundidad o de velocidad. Si el test requiriese idéntico grado de dificultad en las preguntas y los ensayos hubiesen sido realizados con tiempo ilimitado, esta simple prueba nos permitiría considerar la homogeneidad de los elementos en función de la dificultad. También nos discriminaría los resultados cuando exigiésemos diferente grado de dificultad.

Finalmente nos daría un valor para la fidelidad del test, que ahora será algo más diferente al hallado con los métodos de Spearman-Brown, Rulón y Jackson. No podemos olvidar que los elementos integrantes son diversos.

De acuerdo con la tabla III, alcanzamos las siguientes conclusiones:

1. En el test considerado se pueden determinar diferencias individuales

2. En el test estudiado las preguntas son de diferente grado de dificultad.

3. La fidelidad nos produce un valor de r_{tt} = 0,770.

Dejaremos de estudiar las razones por las que este valor es superior al hallado en los métodos anteriores, pero debe reconocerse en este momento la variación producida por el establecimiento de diferentes hipótesis en el caso del método recién estudiado y en los anteriores.

Ahora bien, si al método que acabamos de citar se le pregunta: ¿Son de la misma valía en la determinación de la fidelidad todas las preguntas del test? El método no nos puede contestar. Solamente nos dice si el conjunto de pruebas está formado por preguntas de diferente dificultad.

Ahora bien, a resolver este problema viene otro método que, apoyado también en el analisis de varianza, es de extraordinaria complejidad en el orden de la computación matemática. Este es el *método de Horst.*

En este método se parte igualmente del cuadro de resultados que hemos ofrecido en la tabla I, pero mediante un proceso complejo se determinan no solamente las varianzas propias de cada elemento y del conjunto o total, sino también las covarianzas entre los elementos que integran el test.

Exige por tanto el estudio de las concordancias y diferencias de puntuaciones entre cada par de preguntas.

De acuerdo con nuestro modo de ofrecer estos métodos, presentaremos las matrices que sirven para el estudio del valor de

TABLA IV.—Matriz para el análisis de varianza y covarianzas obtenida de la tabla I.

Elemento	(1) Suma de coincidencias	(2) Covarianzas de elementos total	(3) Varianzas de cada elemento	(4) Suma de las covarianzas con los demás elementos	Indice de selección
1	778	0,000	0,000	0,000	0,000
2	764	0,190	0,028	0,162	0,134
3	738	0,055	0,052	0,003	—0,049
4	721	0,164	0,074	0,090	—0,016
5	709	0,408	0,096	0,312	0,216
6	686	0,355	0,115	0,240	0,125
7	685	0,328	0,115	0,213	0,098
8	665	0,356	0,136	0,220	0,084
9	652	0,573	0,153	0,420	0,267
10	637	0,736	0,170	0,566	0,396
11	614	0,682	0,184	0,498	0,314
12	594	0,710	0,197	0,513	0,316
13	571	0,086	0,197	—0,111	—0,308
14	573	0,711	0,209	0,502	0,293
15	557	0,847	0,219	0,628	0,409
16	519	0,388	0,228	0,160	—0,068
17	528	1,200	0,235	0,965	0,730
18	508	1,227	0,241	0,986	0,745
19	473	0,849	0,245	0,604	0,359
20	427	0,175	0,248	—0,073	—0,321
21	419	0,527	0,250	0,277	—0,027
22	414	0,960	0,250	0,710	0,460
23	393	0,961	0,248	0,713	0,525
24	365	0,772	0,245	0,527	0,282
25	362	0,691	0,245	0,446	0,201
26	374	1,583	0,241	1,342	1,101
27	335	1,098	0,235	0,863	0,628
28	304	0,828	0,228	0,600	0,372
29	303	1,370	0,219	1,151	0,932
30	252	0,559	0,209	0,350	0,141
31	247	0,993	0,197	0,796	0,599
32	229	1,075	0,184	0,891	0,707
33	202	0,913	0,170	0,743	0,573
34	168	0,562	0,153	0,409	0,256
35	167	1,104	0,136	0,968	0,832
36	126	0,564	0,115	0,449	0,334
37	112	0,754	0,096	0,658	0,562
38	84	0,565	.0,074	0,491	0,417
39	65	0,620	0,052	0,568	0,516
40	36	0,405	0,028	0,377	0,349
	17.356	26,944	6,717	20,227	

cada item en el test. Anticipamos que, mediante este sistema, será posible aceptar la totalidad de preguntas de un test o rechazar alguna de ellas. En este segundo caso se rechazarán las preguntas o elementos que reduzcan la verdadera varianza en primer lugar, y las que no contribuyan ampliamente en el total de la verdadera varianza, en segundo. (Tabla IV).

El coeficiente de fidelidad vendria dado por la fórmula

$$r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma^2_{\infty}}{\sigma^2_t} \qquad r_{tt} = \frac{40}{39} \cdot \frac{20{,}227}{26{,}944} = 0{,}771$$

Después de este primer proceso computatorio, hemos obtenido un valor para la fidelidad, realmente idéntico al conseguido mediante el análisis de varianza de Hoyt y el proceso ha sido mucho más complejo.

Pero hemos logrado algo más. Hemos conseguido determinar cuáles son las preguntas que perjudican al test. Es decir, con una simple mirada a la última de las columnas: Indice de selección; advertimos en algunas preguntas o elementos un valor elevado y en otras un valor negativo, nulo o casi nulo.

En principio se debe eliminar toda pregunta con índice de selección negativo. Su contribución a la verdadera varianza es contraproducente. También podríamos eliminar, sin ninguna duda, la pregunta cuyo índice fuese igual a cero. Viene a alargar el test y no favorece la fidelidad. Igual pudiéramos hacer con todo elemento cuyo índice de selección tienda a cero.

Como conclusión práctica diremos que las preguntas que deben ser eliminadas por su valor negativo son las números 3, 4, 13, 16, 20 y 21. También debería serlo la número 1, por ser contestada por todos los escolares y su índice igual a cero. Los elementos 7 y 8 podrán serlo o no debido a que contribyen a la varianza total, pero con muy poca carga.

Ahora bien el proceso de fidelidad no concluye aquí. Una vez eliminadas las preguntas pertinentes se procede a la revisión de las matrices originales. Hecha la nueva estructuración y matriz final, se determina por el mismo proceso el valor definitivo de la fidelidad conseguido entonces con un test de menor número de elementos. En esta última revisión el número de operaciones se reduce considerablemente.

Como compendio de la revisión realizada una vez suprimidos los elementos números 1, 3, 4, 13, 16, 20 y 21 alcanzamos los nuevos cuatro totales correspondientes a las columnas de coincidencia, covarianzas y varianzas.

(1)	(2)	(3)	(4)
10.088	25,699	5,668	20,031

Estos valores nos permiten averiguar el nuevo coeficiente de fidelidad de acuerdo con la fórmula de Kuder-Richardson antes ofrecida:

$$r_{tt} = 0{,}803$$

Si la comparamos con la anterior, el crecimiento en valor absoluto no es excesivo, aunque aumente más la eficiencia de dicho coeficiente. Pero no podemos olvidar que al reducir siete elementos del test anterior la fidelidad del test quedaba o debería quedar disminuída y no obstante es mayor. La fidelidad correspondiente al conjunto formado por cualesquiera de los 33 elementos entresacados por azar debería ser:

$$r_{tt} = \frac{(33:40) \times 0{,}771}{1 - (7:40) \times 0{,}771} = \frac{0{,}636}{0{,}865} = 0{,}735$$

Luego la ganancia real va desde 0,735 a 0,803, con un gran incremento en la probabilidad consiguiente.

Ahora bien el coeficiente de fidelidad hallado antes puede considerarse como una infraestimación o valor mínimo que se cumple cuando la intercorrelación de los elementos de las preguntas fuese la misma. Horst propuso un nuevo procedimiento que por corregir la atenuación producida al no cumplirse la homogeneidad de las preguntas nos da un más elevado coeficiente de correlación.

En la fórmula que ahora exponemos se logra un aumento en los coeficientes de fidelidad sobre los que ya hemos mostrado. Este incremento es proporcionalmente mayor una vez eliminadas las preguntas que producían mayor carga en la varianza del error que en la verdadera varianza.

Los nuevos valores obtenidos con la fórmula que más abajo ofrecemos, son los siguientes:

$$r_{tt\,(40)} = 0{,}791 \qquad r_{tt\,(33)} = 0{,}827$$

Ambos valores son de mayor cuantía a los que se obtuvieron mediante la fórmula general obtenida de la de Kuder-Richardson.

La nueva fórmula de Horst es la siguiente:

$$r_{tt} = \frac{\sigma^2_{\infty}}{\sigma^2_t} \cdot \frac{\sigma^2_m}{\sigma^2_m - \Sigma\,\sigma^2_{il}} \qquad \Sigma\,\sigma^2_{il} = \Sigma_{pq}$$

En esta última fórmula de Horst hemos podido ver cómo se atenúa alguno de los supuestos de la fórmula de Kuder-Richardson. Son varias las fórmulas dadas por estos autores en el estudio de la fidelidad a base del método que lleva el nombre de los autores, a pesar de que ellos mismos le denominaron *método de equivalencia racional.*

Se utiliza mucho tal procedimiento a pesar de que los resultados sufren cierta distorsión cuando el test no es homogéneo. No debemos olvidar que la homogeneidad exige condiciones de equivalencia en dificultad y contenido.

La fórmula que más hemos utilizado es la siguiente:

$$r_{tt} = \frac{n}{n-1}\left(\frac{s^2_t - \sum\limits_{1}^{n} p_i\,q_i}{s^2_t}\right)$$

En la que *n* como en casos anteriores, es el número de elementos de la prueba; s^2_t es la varianza de las puntuaciones conseguidas por los sujetos, p_i es la proporción de aciertos del elemento *i*; $q_i = 1 - p_i$ es la proporción de desaciertos en elemento *i*.

Cuando se opera con pruebas integradas con elementos de idéntica dificultad, puede utilizarse la fórmula simplificada:

$$r_{tt} = \frac{n}{n-1} \cdot \frac{s_t^2 - n\bar{p}\,\bar{q}}{s_t^2}$$

En la que el palor $\bar{p}$ es la media dividida entre el número de preguntas:

$$\bar{q} \text{ igual a } 1 - \bar{p}$$

Aplicada la fórmula primera a nuestro ejemplo obtuvimos el siguiente valor para la fidelidad

$$r_{tt} = 0{,}777$$

Resultado prácticamente igual al hallado en otras fórmulas apoyadas en los mismos principios.

Algo menor es el obtenido al aplicar la segunda $r_{tt} = 0{,}703$.

La infraestimación con esta fórmula de Koder-Richardson es francamente clara. Acarrea un supuesto que suele aparecer cumplido principalmente en los tests de velocidad y por otra parte no puede aplicarse a los tests de rapidez porque muchos examinandos carecen de tiempo para resolver las preguntas finales. En este caso se produciría el fenómeno inverso: sobreestimaríamos el valor de la fidelidad del test. Si perdiese la homogeneidad no por la misma dificultad, sino por la intervención de varios factores en la composición del test, también se produciría la infraestimación que hemos visto intenta salvar Horst con su adaptación del procedimiento de Kuder-Richardson.

En los procedimientos de consistencia es básica la admisión de esta simplificación factorial. En este caso la varianza total puede constituirse a base de la totalización de la varianza y covarianzas conforme la fórmula general:

$$\sigma_t^2 = \Sigma\, \sigma_{ii}^2 + 2\, \Sigma\, r_{ij}\, \sigma_i\, \sigma_j + \sigma_e^2$$

De este modo se incrementa la verdadera varianza cuanto mayores son las covarianzas entre los elementos. Ahora bien, puesto que estas covarianzas están en función de la correlación entre los elementos y las varianzas de cada elemento cuanto mayores sean dichos componentes tanto más elevado será el coeficiente de fidelidad. Por ello, conforme vimos en el método de Horst, alguna de las preguntas podría reducir la fiabilidad. Tal reducción se debía a que la covarianza con los demás elementos era menor que la varianza del mismo elemento.

Fidelidad y generalización

Hemos indicado una serie de métodos para asegurarnos respecto de la confianza que un test nos ofrece para medir exactamente lo que sabemos mide. Ahora bien, una vez conocida dicha fidelidad, ¿hasta qué punto podremos actuar con dicho test y estar seguros de su fidelidad? Conviene que recordemos algunas de las premisas propias de la indagación en el campo de lo experimental. Nosotros podremos generalizar con ciertas limitaciones. No podremos creer que si hemos contrastado la fidelidad de un test respecto de una población con ciertas características, podamos extender la confianza a otra población con caracteres diferentes. A via de ejemplo señalaremos dos grandes grupos diferenciales: estado económico-social y edad. Si en nuestro estudio operamos con sujetos representativos de un determinado sector de la sociedad, pero no de toda ella, no estamos capacitados para asegurar que empleado con sujetos de otro sector el test en cuestión es de la misma confianza. Igual nos acontecería respecto de la edad. Es de sobra conocido que si bien unos tests están bien ajustados y miden con precisión las aptitudes hasta cierta edad, son menos precisos para otras edades. En este caso deberíamos indicar que la generalización del test dependerá de la amplitud de la muestra tomada. Sólo podrá aplicarse con rigor a los sujetos que ofrezcan las características debidamente muestreadas en el estudio de la misma. Claro es que por extensión analógica y con cierto sentido práctico podrá admitirse su validez para los campos limítrofes con el objeto de nuestra indagación, pero con conocimiento de que se puede perder exactitud y por tanto reducir fidelidad.

¿Cómo interpretar por otra parte la posibilidad de que los coeficientes de fidelidad varíen conforme empleemos procedimientos de equivalencia, de consistencia o de estabilidad? El desacuerdo en los resultados no es más que una clara manifestación de que los orígenes de la verdadera varianza son distintos. Por ello podríamos inferir en primer lugar que el test en cuestión estaba constituído por elementos heterogéneos que admitían distinta variabilidad en los errores de medida. El estudio de esta heterogeneidad y la posibilidad de agrupar los elementos podría llevarse a cabo mediante el método de Horst que hemos presentado.

¿Podremos admitir que la aparición de valores altos de fidelidad es suficiente para afirmar su exactitud? Esta pregunta que podríamos contestar de un modo positivo debe ser matizada. Si los métodos que hemos empleado son agrupables en los de consistencia interna, es decir, en los aplicados una sola vez y estudiados por mitades o por covarianzas entre los elementos, entonces hemos de asegurarnos de algo más. Hemos de asegurarnos de si el test en estudio había sido presentado como prueba de velocidad o como prueba de profundidad. Si el test fué presentado como prueba de velocidad e integrada por preguntas que normalmente debe saber el escolar, pero nosotros buscamos su facilidad para resolver muchas en poco tiempo, introducimos un enorme error al utilizar estos procedimientos. Los test de velocidad no toleran ninguno de los procederes que supongan la sola aplicación de un test. Ahora bien, si los test empleados fueron de profundidad o combinaron profundidad con velocidad a base de pruebas de diferente profundidad, pero con tiempo preciso para cada elemento, entonces todos los escolares pudieron contestar a todas las preguntas ya que alcanzaron a intentarlas responder. En este último caso admitiremos la fidelidad con su alto valor si lo tuviere.

José Fernández Huerta
Profesor de Didáctica en la Universidad de Madrid

SUMMARY

In the study on the reliability of the achievement tests we can use several procedures grounded whether on their consistency or on their stability or on both. The author, after remembering the oldest method in the determination of the reliability, presents a series of methods to find out the desired co-efficient. The difference between halves, the sensitivity of Jackson, the variance analysis of Hoyt, Ruder-Richardson and the variance of Horst are offered at the same time that the author points out their merits. A general example to which all these procedures are applied permits to appreciate the different evaluation resulting from some of them.