

ANÁLISIS DE LOS EXÁMENES CONVENCIONALES: LOS RESULTADOS DE UNA OPOSICIÓN A PROFESORES DE E.G.B.

por ANTONIO RODRÍGUEZ DIÉGUEZ

Si bien existe una clara divergencia de posturas ante el tema de la evaluación, lo que realmente está en tela de juicio no es el hecho en sí de la evaluación —sería del campo educativo, de donde únicamente se excluiría esta función— sino que es el mismo método o sistema de aplicación de la evaluación ¹.

Para que puedan ser fidedignos los resultados obtenidos mediante las técnicas de evaluación, es preciso que puedan ser consideradas estas últimas como objetivas, válidas y fiables.

La primera característica, la objetividad en la valoración, quizás una de las mayores dificultades que presentan los tipos de pruebas evaluativas orales y escritas, ha sido ya tratada por Rodríguez Diéguez, J. L., mediante la reducción de los distintos objetivos, a un número concreto y no muy amplio de categorías, dotadas de distintos coeficientes ponderales, de acuerdo con la importancia de los objetivos incluidos en cada una de las categorías ².

La validez y la fiabilidad, características fundamentales de cualquier prueba evaluativa, aplicadas a las pruebas tradicionales orales y escritas, va a ser el objeto de nuestro estudio.

Para realizar la investigación nos basamos en un estudio docimológico realizado por Valin, en un centro de investigación pedagógica libanés, y editado por la UNESCO en París en 1961 ³.

La escasa utilización de las pruebas orales de evaluación, ha condicionado la extensión de la muestra utilizada, si bien es verdad que dicha muestra está dotada de un interés especial por haber sido conseguida con las puntuaciones obtenidas por los candidatos opositores al cuerpo de Profesores de E.G.B. que tuvieron lugar en Alicante en julio del 76.

La naturaleza de esta muestra pone de relieve la problemática actualmente existente sobre la conveniencia o no de las oposiciones como sistema de selección de los candidatos para el ejercicio profesional.

¹ BOMBOIR, A.: *La docimología, problemática de la evaluación*. Edt. Morata, Madrid, 1974, págs. 14-15.

² RODRÍGUEZ DIÉGUEZ, J. L. y Colbs.: *Técnicas de evaluación educativa*. ICE y Dto. de Didáctica de la U. de Valencia, 1979, págs. 29-44.

³ VALIN, E.: *La valeur des examens. Étude docimologique réalisée au Liban*. UNESCO, Paris, 1961.

Se plantean una serie de interrogantes a los que vamos a intentar dar respuesta en lo posible, mediante los resultados obtenidos en la presente investigación:

¿Podría decirse que los resultados de las distintas pruebas se correlacionan entre sí, de modo que nos proporcionen una validez concurrente?

¿Tendremos seguridad de que los candidatos que superan las pruebas propuestas son realmente los más idóneos para el ejercicio de la profesión?

De otra parte, ¿se da una fiabilidad suficiente en la calificación resultante de las pruebas?

¿Están siempre calificadas de igual modo, siguiendo los mismos criterios, por los distintos jueces que intervienen?

Para dar respuesta a las dos primeras preguntas, es decir, correlación entre los resultados de las distintas pruebas y la seguridad de selección de los candidatos más idóneos mediante las mismas, se realizó un estudio de la validez concurrente. La hipótesis que se intentó confirmar es la de que una correlación significativamente alta, entre las distintas pruebas, presupone un valor discriminativo de los candidatos elegidos, es decir, que el tipo de pruebas de que consta la oposición, es el más idóneo para separar a los mejores de los peores candidatos.

La oposición constaba de tres pruebas distintas, dos ejercicios escritos y uno oral, susceptibles de calificación desde cero hasta diez. El primer ejercicio escrito consistía en una prueba de madurez cultural sobre todas las disciplinas cursadas en la carrera. El segundo ejercicio escrito consistía en una prueba de conocimientos específicos, que versaba sobre las áreas de Idioma extranjero, Ciencias Sociales o Matemáticas y Ciencias de la Naturaleza. El ejercicio oral consistía en el desarrollo de tres temas al azar sobre Ciencias de la Educación.

Los resultados de las correlaciones entre las calificaciones del primero, del segundo escrito y del oral es escasa —entre 0,27 y 0,30— lo que nos llevó a deducir una corta validez concurrente. Si analizamos la posible causa, la encontramos en la ausencia de validez de contenidos, ya que como antes expusimos, cada ejercicio versaba sobre áreas diferentes. Otra cosa es la correlación entre cada uno de los tres ejercicios con los resultados finales de las oposiciones, que oscilaba entre 0,64 y 0,69. Incremento de correlación que es explicable, por otra parte, al tener en cuenta que el resultado final no era más que la consecuencia de una función lineal de los resultados de los otros tres ejercicios, más la puntuación de cero a dos concedida por los miembros del tribunal, como valoración de los méritos contraídos por algunos de los opositores que previamente habían ejercido como interinos o contratados en la enseñanza oficial.

Para X_1 , X_2 , X_3 , X_4 , puntuaciones del primer escrito, segundo escrito, ejercicio oral y puntuación final de la oposición respectivamente la tabla de correlaciones obtenidas fue la siguiente:

	X_2	X_3	X_4
X_1	0,3047	0,2778	0,6892
X_2		0,2745	0,6446
X_3			0,6403

Con el fin de probar la significación de los coeficientes de correlación obtenidos, se aplicó el procedimiento descrito por Garrett como «verificación de r frente a la hipótesis nula» para un nivel de significación del 5 %⁴. La correlación que permitía desechar la hipótesis nula, tenía que ascender a 0,266. Como las correlaciones halladas estaban precisamente alrededor de 0,27, el rechazo de la hipótesis nula se daba, pero resultaba demasiado ajustado a un nivel de confianza del 0,05. El coeficiente de correlación era muy débilmente significativo; se daba, por tanto, una muy corta validez concurrente⁵.

Para responder a la pregunta que nos hicimos sobre la fiabilidad en la calificación de las pruebas, se trató de buscar la puntuación estimada (P_e) que se podía esperar de un sujeto si el instrumento de evaluación fuese idealmente fiable. Hallado el intervalo confidencial de puntuaciones en el cual un determinado alumno estaría incluido a un nivel de confianza de, al menos, 5 %, se lo añadimos en más o en menos a la puntuación estimada.

Este método se aplicó a las calificaciones obtenidas por cada uno de los opositores en la prueba escrita, que constaba de dos partes —problema matemático y tema de madurez general— y en el ejercicio oral.

Del problema matemático evaluado por dos profesores se obtuvieron los siguientes datos de interés.

$$\begin{array}{ll} N = 86 & \sigma = 2,108 \\ \bar{X} = 2,539 & r = 0,980 \end{array}$$

Aplicando a estos datos la fórmula de la puntuación estimada y sumando y restando los valores obtenidos en el intervalo confidencial al nivel del 5 %, obtuvimos el resultado siguiente:

$$2,01 \pm 0,576$$

Como a todos los opositores se les añadió tres puntos en cada una de las partes del ejercicio escrito, según el criterio del tribunal, el intervalo confidencial quedaba establecido en:

$$5,01 \pm 0,576 = \begin{cases} 5,586 \\ 4,434 \end{cases}$$

⁴ GARRETT, H. E.: *Estadística en Psicología y Educación*. Paidós, Buenos Aires, 1976, págs. 228-231.

⁵ Datos parciales de esta investigación, fueron utilizados, con mi autorización, por J. L. RODRÍGUEZ DIÉGUEZ en su obra: *Didáctica General. Objetivos y Evaluación*. Edt. Cincel, Madrid, 1980.

Así pues, un opositor que hubiese sido rechazado con un 4,45 podría haber obtenido un 5 y así superar el ejercicio en no menos del 5 % de los casos.

El resultado en el tema de madurez general fue aún más significativo. Para los datos:

$$\begin{aligned} N &= 77 & \sigma &= 1,246 \\ \bar{X} &= 2,830 \end{aligned}$$

Al ser calificado por cuatro jueces, fue necesario averiguar el coeficiente de correlación promedio según la fórmula propuesta por Garrett⁶. El resultado obtenido en dicho coeficiente fue:

$$r \text{ promed.} = 0,65$$

Añadiendo a la puntuación estimada los tres puntos concedidos según acuerdo del tribunal, y sumando y restando el intervalo confidencial a la misma el resultado fue:

$$5,290 \pm 1,444 = \begin{cases} 6,734 \\ 3,846 \end{cases}$$

Así, un opositor que hubiera obtenido un 3,85, podría haber obtenido 5 en no menos del 5 % de los casos y así superar el ejercicio.

Para la prueba oral, en la que los candidatos habían de desarrollar tres temas cada uno, y éstos eran calificados por seis miembros del tribunal, después de aplicar la fórmula anteriormente citada por Garret para obtener la correlación promedio, los datos estadísticos obtenidos fueron los siguientes:

$$\begin{aligned} N &= 153 \text{ (51 candidatos por 3 ejercicios cada uno)} \\ \bar{X} &= 4,846 \\ \sigma &= 1,688 \\ r &= 0,79 \end{aligned}$$

Una vez obtenida la puntuación estimada y el intervalo confidencial el resultado en la prueba oral fue:

$$4,961 \pm 1,516 = \begin{cases} 6,477 \\ 3,445 \end{cases}$$

Así, un ejercicio que se hubiera calificado con 3,50, en no menos del 5 % de los casos, podría haber sido calificado con 5 puntos, mínimo para ser aceptado.

Conviene observar algunas consecuencias significativas.

⁶ GARRETT, H. E.: *Op. cit.*, págs. 201-202.

La mayor amplitud del intervalo corresponde al ejercicio oral —más de tres puntos—, poniendo de manifiesto la dificultad de valorar con objetividad una prueba oral. La menor amplitud corresponde sin embargo a la prueba matemática del escrito, para cuya corrección se dispone sin duda de unos criterios más fijos y objetivos de evaluación.

Las puntuaciones intermedias comprendidas entre la puntuación estimada más menos el intervalo confidencial son significativamente iguales. Así, alumnos definitivamente rechazados, podrían haber superado el ejercicio oposición, en muchos casos con una probabilidad superior al 5 %.

ANÁLISIS ESPECIAL DE LA PRUEBA ORAL

El mayor problema que presentan las pruebas orales es el de la calificación objetiva, y esto por la misma naturaleza de la prueba, puesto que mientras la prueba escrita suele ser la misma para un número elevado de candidatos, susceptible por tanto de tener preparado un baremo de calificaciones lo más objetivo posible, en la prueba oral, por su característica de repentización que suele acompañarle, eligiendo temas al azar, no es posible tener previsto un baremo fijo que compute con objetividad el desarrollo de cada uno de los temas. La prueba escrita, por otra parte, es susceptible de repaso y revisión, mientras que el oral no permite una nueva revisión.

Un intento de afrontar la falta de fiabilidad de la calificación otorgada en la prueba oral de las oposiciones era el hecho de que para dicha prueba el opositor era valorado por seis jueces en cada una de las tres pruebas de que se componía.

Datos estadísticos de cada uno de los seis jueces:

Juez n.º 1:

$$\begin{array}{ll} \bar{X} = 4,9891 & \Sigma f = 231 \\ \sigma = 1,7253 & \Sigma X = 1152,5 \end{array}$$

Juez n.º 2:

$$\begin{array}{ll} \bar{X} = 4,7337 & \Sigma f = 231 \\ \sigma = 1,8240 & \Sigma X = 1093,5 \end{array}$$

Juez n.º 3:

$$\begin{array}{ll} \bar{X} = 4,6558 & \Sigma f = 231 \\ \sigma = 1,9684 & \Sigma X = 1075,5 \end{array}$$

Juez n.º 4:

$$\begin{array}{ll} \bar{X} = 5,2186 & \Sigma f = 231 \\ \sigma = 1,4755 & \Sigma X = 1202,95 \end{array}$$

Juez n.º 5:

$$\begin{aligned}\bar{X} &= 5,1556 & \Sigma f &= 231 \\ \sigma &= 1,7129 & \Sigma X &= 1190,77\end{aligned}$$

Juez n.º 6:

$$\begin{aligned}\bar{X} &= 5,0487 & \Sigma f &= 231 \\ \sigma &= 1,5515 & \Sigma X &= 1166,25\end{aligned}$$

¿Hasta qué punto las diferencias de valores observados son significativas?

Mediante el análisis de varianza determinamos la significación de diferencias entre las medias de las calificaciones totales, otorgadas por cada uno de los jueces. Tratamos de determinar, por tanto, si las diferencias entre las medias son reales o si hay que aceptar la hipótesis nula, según la cual, no hay diferencias entre las medias, y si la hay, es accidental y sin importancia debida al azar.

Resumen del análisis de varianza.

Fuente de varianza	Grados de libertad	Suma de cuadrados	Cuadrados medios (V)	σ
Inter	5	59,430	11,8860	
Intra	1380	3983,846	2,8868	1,699
TOTAL	1385	4043,276		

$$F = \frac{11,8860}{2,8868} = 4,1173$$

Según la tabla F para g.l. 5 y para g.l. 1000 ⁷

F al nivel de confianza de 0,05 = 2,22

Para estudiar la significación de diferencias entre las medias obtenidas por cada juez, seguimos a Garrett en el análisis del error estándar de la diferencia entre dos medias cualesquiera en análisis de varianza ⁸.

Después de consultar la tabla —D— para la prueba «t» para determinar la significación de las estadísticas con 1000 grados de libertad ⁹, llegamos a la conclusión que toda diferencia entre las medias de las puntuaciones de cada uno de los jueces, superior a 0,3096 es significativamente diferente con nivel de confianza

⁷ GARRETT, H. E.: *Op. cit.*, pág. 492.

⁸ GARRETT, H. E.: *Op. cit.*, pág. 317.

⁹ GARRETT, H. E.: *Op. cit.*, pág. 487.

del 5 %. Toda diferencia entre medias, superior a 0,4076, es significativamente diferente al nivel de confianza del 1 %.

Así la significación de diferencias la conseguimos comparando las medias de cada juez con cada una de las medias restantes.

Se realizó la estimación a nivel de 0,1 y de 0,5. De los resultados de este último daremos cuenta, ya que el 0,5 es considerado satisfactorio para este tipo de investigación.

No son significativas al nivel de confianza del 0,5 las siguientes medias:

$$\begin{array}{ccc} \bar{X}_1 \text{ y } \bar{X}_2 & \bar{X}_6 \text{ y } \bar{X}_1 & \bar{X}_4 \text{ y } \bar{X}_6 \\ \bar{X}_4 \text{ y } \bar{X}_1 & \bar{X}_3 \text{ y } \bar{X}_3 & \bar{X}_5 \text{ y } \bar{X}_6 \\ \bar{X}_5 \text{ y } \bar{X}_1 & \bar{X}_4 \text{ y } \bar{X}_5 & \end{array}$$

Son significativas al nivel de confianza del 0,5 las diferencias entre las siguientes medias:

$$\begin{array}{ccc} \bar{X}_1 \text{ y } \bar{X}_3 & \bar{X}_6 \text{ y } \bar{X}_2 & \bar{X}_5 \text{ y } \bar{X}_3 \\ \bar{X}_4 \text{ y } \bar{X}_2 & \bar{X}_4 \text{ y } \bar{X}_3 & \bar{X}_6 \text{ y } \bar{X}_3 \\ \bar{X}_5 \text{ y } \bar{X}_2 & & \end{array}$$

Por tanto, de las quince combinaciones posibles de diferencias entre medias, se han obtenido ocho combinaciones en las que no se da una diferencia significativa, y siete que se pueden considerar significativamente diferentes. El 46,66 % de las medias están afectadas, por tanto, de diferencias importantes, no atribuibles al azar.

CONCLUSIÓN

El resultado de la investigación es un testimonio de la muy relativa validez y de la práctica ausencia de fiabilidad del tipo de exámenes convencionales; poniendo en tela de juicio la estructura actual del sistema de oposiciones, para seleccionar a los candidatos más idóneos.

De alguna manera queda así justificada la contestación que hoy se está dando, cada vez con más rigor, de amplios sectores de candidatos al ingreso en la enseñanza oficial.

A la vista de este estudio, con la limitación que supone lo reducido de la muestra, habría que someter a crítica si el tipo de pruebas son las más idóneas para seleccionar a los opositores, y si el sistema de calificación es lo suficientemente fiable como para atribuirle el poder decisorio de aceptar o rechazar a ciertos candidatos.

Hubiera sido de gran interés realizar un estudio de la validez predictiva, pero se salía por completo de este objetivo y de nuestras posibilidades, por la dificultad

que entraña un seguimiento longitudinal del quehacer profesional hasta hoy, de aquellos opositores que obtuvieron plaza en el cuerpo de profesores de E.G.B.

El problema reside en hallar un sistema de selección de candidatos más válido y fiable que el vigente, que pudiera suplirle con eficacia, viéndose libre de posibles variables extrañas de orden político, social o cualquier otra.

La solución propuesta por Valin ¹⁰ en su estudio docimológico anteriormente citado pasaba por una tripartición racional de los resultados en: candidatos claramente aceptados, claramente rechazados y dudosos. Sería a estos dudosos a quienes debería sometérselos a otra sesión, que equilibrase y rectificase en lo posible los errores de diferencias significativas de puntuaciones.

ANTONIO RODRÍGUEZ DIÉGUEZ

Profesor de la Sección de Ciencias de la Educación
Facultad de Filosofía y Ciencias de la Educación

GONZALO BILBAO, 5
SEVILLA

¹⁰ VALIN, E.: *Op. cit.*