

PUNTUACIONES TIPICAS EN LA COMPOSICION DE NOTAS

El creciente poder de los exámenes, que con su urgencia resquebraja a los débiles y tambalea a los fuertes, no es reducible por mucho que se discuta. A las críticas respecto de fortuidad, memorismo, servilismo intelectual, evidencia de justicia..., se responde con una gama completa de matices (1) en tan exuberante variedad que se alcanza o intenta alcanzar la coordinación de pruebas subjetivas y objetivas. Mas toda esta proliferación «docimológica» de tipo analítico ha originado nuevos problemas: ¿Cuántas serán las subdivisiones de la batería de pruebas? ¿Cómo compararemos las puntuaciones obtenidas con diversos tests mentales o instructivos? ¿Cuándo legitimaremos la composición de notas en una sola, sean o no comparables los resultados?... Las interrogaciones fluyen con gran diversidad y se estrellan sucesivamente ante los procedimientos vulgares. El ánimo del examinador inquieto, preocupado, estudioso, se turba ante tantas dificultades y pierde el sosiego. ¿Bastará examinar de acuerdo con lo legislado si nuestra conciencia de examinador queda insatisfecha por las normas emanadas de la superioridad? Una fina estimación de los valores inherentes a las pruebas, propia de examinadores con sentido psicopedagógico, es causa de un sentimiento de justicia incompleta. Anhelamos eliminar este sentimiento muy próximo al de culpabilidad. En todo caso hemos de procurar minimizar los efectos perniciosos de este sentir. Pretendemos, en suma, diluir las principales objeciones presentadas respecto de la comparabilidad y composición de las notas obtenidas.

En el campo de los tests mentales tropezamos contra grandes obstáculos al comparar resultados obtenidos con pruebas distintas. ¿Quién afirmaría que el cociente intelectual obtenido por medio de la prueba A es directamente comparable con el dado en la prueba B? ¿Quién mantendría, desde otro punto de vista, que el centil ocupado por un sujeto en el test C es equivalente al obtenido en el test D? ¿Quién defendería que, supuesta la fidelidad máxima en cada test, el cociente intelectual de un sujeto es inalterable? Vemos como en el campo de los tests mentales la comparabilidad está sujeta a severas restricciones.

Restricciones que se incrementan con agudos caracteres cuando nos

(1) Véase: «Formas de examen del rendimiento escolar».—*Revista Española de Pedagogía*, núm. 39. 1952, págs. 377-390.

Evidencia y tipos de justicia examinadora.—*Bordón*, núm. 31, 1952, págs. 349-354.

Técnica de los exámenes universitarios.—*Revista de Educación*, núm. 14, 1953, páginas 249-251.

trasladamos al terreno didáctico. No sólo porque los tests instructivos han logrado un menor grado de desenvolvimiento y porque su fiabilidad no ha alcanzado el alto rango de muchos tests mentales, sino porque el número de situaciones complejas aumenta. ¿Quién sería capaz de obtener un test de cualquier materia cuya correlación con los factores de grupo más importantes fuese realmente anulable? ¿Qué funciones ha de examinar un test instructivo de alguna asignatura? Como puede advertirse, cuando se presentan las características generales o específicas (2) de pruebas instructivas, el problema es complejo. Tan complicado, que no está definitivamente resuelto aunque se haya avanzado en alguno de los sentidos resolutivos.

La visión de complejidad de este u otro problemas es un índice de potencia intelectual, al igual que la negativa a tan amplia problemática es señal de impotencia intelectual. Mas no todos los problemas pueden ser resueltos satisfactoriamente. ¿Será esta cuestión un problema de pura aporía? Si lo fuese entenderíamos que sería más conveniente abandonar los problemas técnicos derivados que insistir sobre ellos. Por no parecernos aporético hemos de atacar fundamentos y técnicas, pero ¿qué haremos para suavizar las dificultades?

Ningún profesor discute la necesidad que como juez se le señalen las vías de solución. En todo momento se procura el logro de la nota conjunta, bien por el docente en su asignatura, bien por un grupo de profesores respecto de una o varias materias.

Si reflexionamos un instante comprendemos que la reducción a una de todas las notas es ilegítima cuando dichas puntuaciones son incomparables. Sentimos la necesidad de reducir las notas a una sola, pero advertimos la subordinación a ciertos supuestos: las puntuaciones consideradas han de ser primero comparables, luego componibles.

En cuanto pedagogos mantenemos que toda nota compuesta sólo es posible cuando lo son las notas simples que la integran, siempre que sean previamente comparables. Las soluciones que presentemos han de ser virtualmente prepotentes, han de estar preñadas de aciertos y potencialidades. El saber nuestro en tanto pedagogo procurará desvelar los fundamentos y desenmarañar las técnicas que permitan eliminar los errores comparativos. No sólo han de ser evitados los errores iniciales, sino también los intermedios y finales.

Una de las fuentes más caudalosas de error se apoya en las comparaciones injustificadas correspondientes a lo que pronto calificaremos como primer paso en la comparabilidad. Comparaciones injustificadas son las que se realizan entre notas o puntuaciones de situaciones incompa-

(2) Véase: Elaboración empírica y científica de las diferentes pruebas de instrucción.—*Bordón*, núm. 21, 1951, págs. 241-245.

Determinación del rendimiento escolar en ciencias naturales elementales.—*Bordón*, número 34, 1953, págs. 157-165.

Evaluación del progreso escolar en ciencias sociales.—*Bordón*, núm. 39, 1953, páginas 719-728.

rables, o las que fundamentan toda la labor en las notas originales.

Dado que entre nosotros no es corriente aplicar el criterio de puntuaciones «verdaderas», apoyado en la fidelidad de las pruebas, cabe señalar que la mayoría de las comparaciones realizadas en el campo de la docencia pertenecer a este grupo denominado de comparaciones injustificadas. ¿Qué razones podríamos encontrar para explicar tal situación? La primera pudiera ser: sencillez operatoria; la segunda: desconocimiento de procedimientos mejores. ¿Es justo, preguntamos, que por facilidad o ignorancia operemos con las notas «brutas»? ¿Compensa la sencillez el número de errores que aparecerán o el fracaso interpretativo?

Pensemos en una situación práctica. Examinado un escolar con tres pruebas diferentes ha obtenido en las tres la puntuación 7. ¿Acaso estos números siete correspondientes a aptitudes diversas pueden tener el mismo significado? Es cierto que si en algunos casos pudieran poseer significado idéntico, en otros no. Así en una de las pruebas siete pudiera ser el máximo alcanzado por el colectivo, mientras que en otro podría ser el mínimo, y en el tercero, constituir el valor central de todos los valores. La equivalencia significativa exige otros conceptos que puedan transformar las puntuaciones «brutas» en puntuaciones definitivas.

Un primer paso, hace poco anunciado, en el campo de la comparabilidad nos viene dado por esta calificación corriente injustificada, que desmenuzaremos algo más para hacer patentes sus ligerezas. El escolar preguntado recibe puntuaciones que variarán desde 0 hasta 4, 10, 15..., según el sistema adoptado. Estas calificaciones van acotadas en muchos casos por indicaciones precisas, que señalan los límites de aprobado, notable, sobresaliente... Si es de base 10 suelen presentarse los límites: 0 a 4,99, de 5 a 6,49..., para determinar la calificación: suspenso, aprobado... A muchos les ha parecido que esta distribución de notas y calificaciones perjudica al alumno porque se ofrecen cinco notas o fracción para el suspenso, dos notas a lo sumo para aprobado, una o dos notas para el notable... No es justa la censura. El proceso lógico inicial consiste en una especie de clasificación dicotómica: de 0 a 5 son puntuados los ineptos, de 5 en adelante los aptos. Pero mientras para los ineptos no se subdivide y matiza su deficiencia, para los aptos se admite diferente gradación. Hay cuatro tipos de aptitud: aprobado, notable, sobresaliente y matrícula de honor. Si hubiesen mantenido el criterio dicotómico anterior, lógicamente explicable, las puntuaciones límites para el paso de una nota a otra serían: 7,50, 8,75 y 9,37, pero para aumentar el porcentaje de alumnos con buenas notas estos topes han sido bastante reducidos.

Desde punto de vista crítico, este procedimiento es en todo similar al obtenido mediante exámenes objetivos de cien preguntas y puntuación de los aciertos, sea o no corregida la fortuidad. El resultado puede admitirse como porcentaje o bien como subdivisión de la nota de 0 a 10.

A porcentaje podrían ser reducidas todas las puntuaciones obtenidas en cualquier examen, independiente del número de preguntas, por una

sencilla razón entre el número de aciertos y el total de preguntas. Esta ligera transformación no produce valores más comparables que la puntuación de 0 al 10, ni que la anotación del resultado de las cien pruebas objetivas.

Pero todos estos sistemas son desechados porque se centran en un supuesto: la puntuación como valor absoluto. ¿Sería posible que creyésemos en nuestra puntuación como un valor indiscutible e invariable? ¿Acaso nuestra nota de 7 representa un juicio constante? Pertenece al dominio público que cuando el examen es subjetivo pocos jueces coinciden con sus mismas notas (3), y si es objetivo, tests de instrucción, la fiabilidad nunca alcanza el valor 1, exigible en hipótesis de nota fija. ¿Quién, a estas alturas, imaginaria que mensuración psicopedagógica y física coinciden? ¿Es tan fácilmente postulable el cero pedagógico como el matemático? En los primeros hay más gradación o estimación que en los segundos. Mientras que en la medida psicopedagógica la variabilidad pertenece a su misma constitución y es radicalmente ineliminable, aunque sea muy reducible, en la física es admisible se considere que el error tiende a cero.

Al no admitir el valor absoluto de las puntuaciones, que acarrearían la dificultad de la determinación del verdadero valor cero de la escala, son necesarias otras medidas o valores que relativicen las puntuaciones. Al igual que la puntuación siete mentada al comienzo, podríamos decir que las puntuaciones 25 ó 75 obtenidas con tests instructivos no son equiparables con otros 25 ó 75 en prueba diferente, mientras no cumplan otras condiciones. La condición básica es que dichas puntuaciones estén emplazadas en la misma posición dentro de la distribución adoptada como propia de dicho examen.

Precisamente a resolver esta primera dificultad viene un procedimiento superior a los anteriores: *el centilaje*. El centilaje es un refinamiento del último proceso consistente en una doble transformación: de puntuación a orden y de orden simple a orden porcentual. La calificación correspondiente a cada sujeto no vendrá dada por los puntos obtenidos en dicha prueba, sino por el orden porcentual alcanzado en la segunda transformación. Desaparece la inconveniencia de que las puntuaciones obtenidas pudieran corresponder en un caso a máximo y en otro a mínimo. El máximo será siempre el mismo: 100, es decir, el sujeto que más puntos ha alcanzado, y el mínimo: 0, es decir, el escolar de menor puntuación de todos los examinados.

Ante un caso práctico resolveríamos de la siguiente manera: no afirmaríamos que el sujeto A obtuvo una nota de 30 puntos en una asignatura y 60 puntos en otra, sino que diríamos el sujeto A ocupa el centil 48 en tal materia y el centil 52 en la otra, si después de las transformaciones antedichas son esos los centiles en que encajan las puntua-

(3) FERNÁNDEZ HUERTA, J.: *Las pruebas objetivas en la escuela primaria*. C. S. I. C. Madrid, 1950.

ciones 30 y 60 respectivamente. Las puntuaciones se refieren de este modo a posiciones, a porcentaje posicional. Deja de preocupar la facilidad o dificultad de los exámenes sin que produzca deterioros la menor capacidad discriminativa de una prueba.

Precisamente algunas de las objeciones más flojas contra este proceder se centran en su sabor posicional. No hay puntos, sino posiciones. ¿No despreciaremos con esto una información tan precisa como la puntuación mediante test instructivo? La defensa contra esta dificultad es sencilla. Se apoyará en la consideración del valor de los exámenes en cuanto notificable o cuantificable, o en cuanto calificable (4). Más seria es la objeción que se apoya en el estudio de las distribuciones normales.

El examen de las distribuciones normales hace patente la diferencia entre distancia percentilada y verdadera distancia en puntuaciones. Así se puede comprobar, de acuerdo con la distribución normal, que la diferencia 1 por 100 existente entre el centil 51 y el 52 es en realidad mucho menor que el 1 por 100 entre los centiles 98 y 99. En tanto las diferencias de centilajes son diversas, quiere decir que a cada lado de la media las unidades verdaderas varían de modo constante cuando las referimos a las unidades de centilaje. Si las unidades varían, admitimos lógicamente que estamos operando con medidas heterogéneas. ¿Sería justo que la media aritmética de las puntuaciones traducibles en los centiles 48, 50, 52, verificada con dichos centiles e igual a 50, fuese técnicamente igual a la conseguible con los centiles 40, 42, 68? Desde perspectiva puramente centilada no cabría objeción: las medias serían iguales aunque no lo fuesen las variabilidades. Traducido en sistema normal el segundo sujeto es ligeramente superior al primero. Luego debe variarse el procedimiento.

Todas las dificultades recién presentadas se vencen si calificamos a los sujetos bajo la base de las *puntuaciones típicas o tipificadas*. Se denomina puntuación típica o tipificada a la resultante de transformar el valor medio en cero y la desviación típica en 1. La media se convertirá en cero cuando restemos a todas las puntuaciones el valor medio, ya que:

$$M' = \frac{\Sigma(X - \bar{X})}{N} = \frac{\Sigma X - N\bar{X}}{N} = \frac{\Sigma X}{N} - \bar{X} = \bar{X} - \bar{X} = 0$$

La desviación típica se iguala a uno al dividir todas las puntuaciones entre la sigma de la muestra. Por la transformación la media será: $\frac{\bar{X}}{S}$

(4) Véase: ¿Conviene la neta media simple en los exámenes conjuntos de enseñanza secundaria? *Revista de Educación*, núm. 23. Julio-agosto (en prensa).

Luego:

$$S' = \sqrt{\frac{\sum \left(\frac{X}{S} - \frac{\bar{X}}{S} \right)^2}{N-1}} = \sqrt{\frac{\sum (X - \bar{X})^2}{(N-1) S^2}} = \frac{1}{S} \sqrt{\frac{\sum (X - \bar{X})^2}{N-1}} = \frac{1}{S} \cdot S = 1$$

Ofrecen las puntuaciones típicas la garantía de que cada punto obtenido por el sujeto influye sobre el final con idéntica intensidad. Igual se incrementa la puntuación tipificada al paso de 9 a 10, que de 49 a 50, ó 79 a 80. Lo que sí será algo distinto es el significado probabilístico de dichas puntuaciones típicas.

El proceso de obtención es realmente fácil. En su forma más simple se opera del siguiente modo: $Z = (X - \bar{X}) (1 : S)$. La labor se facilita si construimos una tabla con el producto de los números naturales por el inverso sigma.

En la tabla I ofrecemos su ejemplo cuando $M = 48$ y $s = 12$.

TABLA I.-De productos para obtener puntuaciones típicas

Media = \bar{X} 48		Desviación típica = $S = 12$		Inverso = $1 : S = 0,0833$	
$ X - \bar{X} $	Puntuación típica	$ X - \bar{X} $	Puntuación típica	$ X - \bar{X} $	Puntuación típica
1	0,0833	7	0,5833	40	3,3333
2	0,1667	8	0,6667	50	4,1667
3	0,2500	9	0,7500	60	5,0000
4	0,3333	10	0,8333	70	5,8333
5	0,4167	20	1,6667	80	6,6667
6	0,5000	30	2,5000	90	7,5000

Si utilizamos esta tabla de productos de números inversos, las operaciones son realmente sencillas. Así las z correspondiente a las puntuaciones 63 y 20 serán, respectivamente:

$$Z_{63} = (63 - 48) \cdot 0,0833 = 15 \cdot 0,0833 = (10 + 5) \cdot 0,0833 = 0,8333 + 0,4167 = 1,25$$

$$Z_{20} = (20 - 48) \cdot 0,0833 = -28 \cdot 0,0833 = -(20 + 8) \cdot 0,0833 = -(1,6667 + 0,6667) = -2,3333 \approx -3,33$$

Tanto en este ejemplo como en lo razonado se puede advertir la tendencia relativizante de las puntuaciones típicas y la característica de

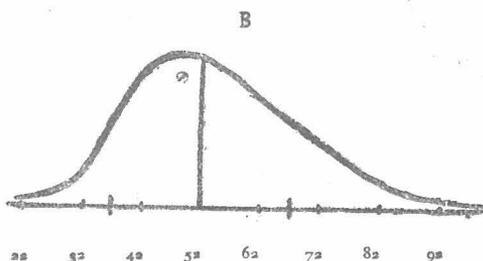
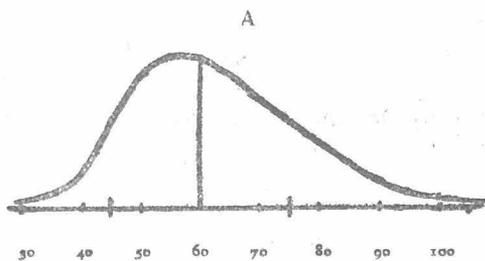
que cada punto vale lo mismo esté en la posición que esté. El valor correspondiente a cada punto de nuestro ejemplo es de 0,0833 sigma.

Se ha adelantado respecto de los sistemas anteriores, si nos atenemos al concepto de homogeneidad de las puntuaciones, ya que todas las medidas y desviaciones típicas que entran en juego son las mismas, pero subsisten algunas limitaciones.

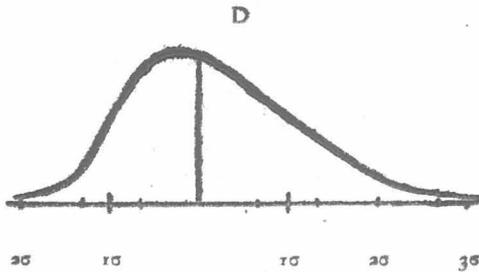
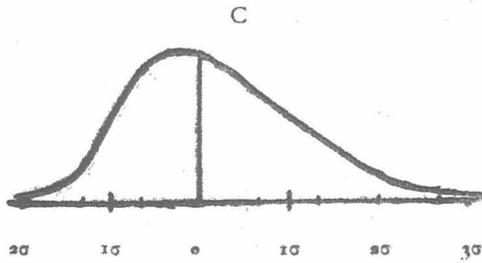
En primer lugar, al tipicar las notas no hemos logrado que las distribuciones adquieran la misma forma. La forma de las distribuciones, una vez definidas por los dos primeros momentos, viene dada por los tercer y cuarto momentos. Cuando cada uno de estos dos últimos momentos sea estadísticamente indiferenciable, se admite que la forma es equivalente. No olvidar que actuamos ya bajo la igualdad de media y sigma.

La demostración más sencilla de la desigualdad de forma, lo que arrastra tras sí el concepto de heterogeneidad, se puede verificar mediante las siguientes curvas de distribución.

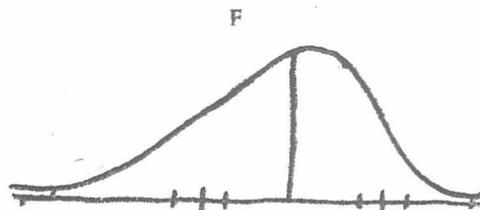
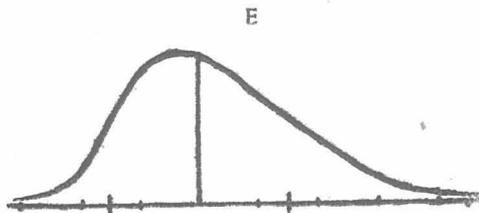
Supongamos dos materias, A y B, en las cuales las puntuaciones se han distribuido de la siguiente manera:



En este caso, aunque media y desviación típica sean diferentes, la forma de las dos distribuciones es idéntica. Con el proceso de tipificación anterior obtendríamos estas dos nuevas curvas, también idénticas en forma, pero con estadios de homogeneidad más perfecto al conseguir que las medias fuesen iguales a 0 y las desviaciones típicas a 1.



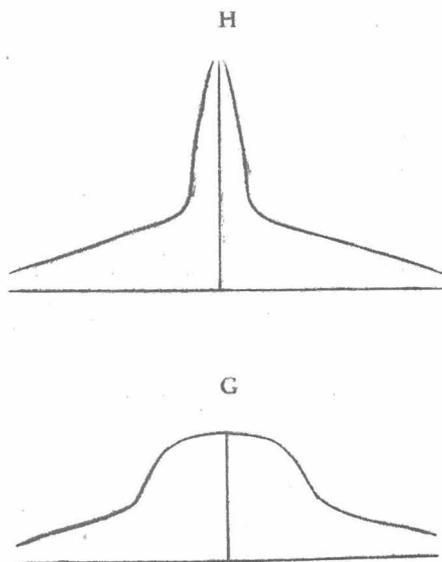
Mas si nos encontrásemos con que las puntuaciones se distribuyen de acuerdo con estas dos representaciones gráficas,



alcanzaríamos la conclusión de que si esta diferenciación formal es significativa, no merecería la pena tipificar las puntuaciones. La heterogeneidad salta a la vista. La comparabilidad de las notas sería radical-

mente injustificada (5) al comprobar el nivel de significación de la discrepancia formal.

Pero el ejemplo mostrado sólo se refería al tercer momento, a la asimetría, y conforme hemos dicho la forma se precisa por los tercer y cuarto momentos. Un nuevo ejemplo gráfico puede mostrarnos dos distribuciones que concordes en los tres primeros momentos difieran radicalmente en el cuarto. Veámoslo.



Ante estas dos distribuciones, concluimos como anteriormente. Las notas de una y otra materia no son componibles.

Cuando quiere suprimirse la comparación formal o cuando de hecho se suprime estamos bajo el efecto de una hipótesis más o menos explícita: cuando media y sigma son iguales, las dos distribuciones son también iguales. Supuesto que informa realmente todo el proceder técnico en la composición de notas tipificadas. Hipótesis más justificada cuando son los mismos sujetos los que se comparan en el conjunto de todas las notas porque entonces cabe aceptar la uniformidad muestral.

Si para continuar el proceso queremos hacer patentes las diferencias entre las puntuaciones tipificadas y las puntuaciones brutas, hemos de recurrir a una tabla en la que presentemos casos supuestos. Para mayor

(5) Véase: El criterio intrínseco en la hipótesis de normalidad dentro de la experimentación didáctico-pedagógica.—*Revista Española de Pedagogía*, núm. 44, 1953, páginas 517-527.

sencillez operatoria elegimos los valores medio y desviación típica, de modo que sean enteros y caigan dentro del intervalo de confianza.

TABLA II

Núm. de sujetos $N_1=N_2=N_3=$ $=N_4=N_5=$	Momentos muestrales		Puntuaciones brutas de sujetos			Puntuaciones típicas (Z) de sujetos			Diferencias $ Z-M_x $		
	\bar{X}	S	1.º	2.º	3.º	1.º	2.º	3.º	1.º	2.º	3.º
Materias											
Matemáticas	60	30	70	40	60	+0,33	-0,66	0,00	0,51	1,33	0,63
Ciencias ...	32	5	20	50	42	-2,40	+3,60	+2,00	2,22	2,93	1,37
Lenguaje ..	85	15	85	78	93	0,00	-0,47	+0,53	0,18	1,14	0,10
Historia	47	10	53	58	66	+0,60	+1,10	+1,90	0,78	0,43	1,27
Religión ...	53	20	64	49	27	+0,55	-0,20	-1,30	0,73	0,87	1,93
TOTALES ..	277		292	275	288	-0,92	+3,37	+3,13	4,42	6,70	5,30
Medias ..	55,4		58,4	55,0	57,6	-0,18	+0,67	+0,63	0,88	1,34	1,06

Según la tabla II, y de acuerdo con las puntuaciones brutas, el mejor sujeto, bajo el supuesto de idéntico índice de ponderación, sería el numerado con el 1, seguido del 3 y 2. Si operamos con puntuaciones típicas, el orden quedaría invertido y sería así: 2, 3, 1. ¿Cómo podría intentarse explicarse esto que podríamos llamar anomalía? Conviene que aprendamos a fijarnos en las tablas numéricas. Posemos la atención sobre la columna de las s. Con esta columna delante advertimos que un punto en ciencias vale lo que dos de historia, tres de lenguaje, cuatro de religión y seis de matemáticas. ¿Por qué se nos preguntará se justifican tales valores relativos? No debemos olvidar lo que representa la desviación típica. Por otra parte, la lógica elemental nos dice que aumentar cinco puntos en ciencias es tan difícil como aumentar treinta en matemáticas, luego aumentar un punto en ciencias es tan dificultoso como seis en matemáticas. Tendrá así el mismo mérito subir dos puntos en ciencias como seis en lenguaje. La valía de cada incremento de un punto en las diferentes materias está en razón inversa a las desviaciones típicas. Las razones para tales discrepancias no son del caso y aunque debieran ser reducidas conseguimos de este modo un ejemplo mucho más claro.

¿En qué radica la superioridad del sujeto primero respecto del segundo en puntuaciones brutas? En que le supera en 30 puntos en matemáticas, 15 en religión y 7 en lenguaje (total, 52), mientras que es superado

con 30 en ciencias y 5 en historia (total, 35). Mas si multiplicamos por las equivalencias anteriores, obtendremos para el primero (haremos la referencia sobre ciencias) $5 + 3,75 + 2,33 = 11,08$, y para el segundo, $30 + 2,5 = 32,5$. Se justifica así lógicamente lo que estadísticamente no ofrece dificultad. El segundo alumno es el mejor de los tres.

Algunos autores con perspectiva diagnóstica proponen se considere la intravariabilidad (variabilidad típica en puntuaciones tipificadas). Con tal procedimiento es fácil advertir la ineptitud o aptitud globales junto al florecimiento de alguna aptitud o defectuosidad de otra. Pero este proceder, presentaba en la tabla II con los resultados de 0,88, 1,34 y 1,06 para los tres sujetos, no se ha extendido lo suficiente porque se apoya, en dos supuestos no comprobados: 1) La verdadera variabilidad de cada aptitud es la misma. 2) La fiabilidad aceptable es estadísticamente igual.

Si regresamos a las puntuaciones tipificadas hemos de recoger las dos principales críticas que, bajo el supuesto de identidad formal, se le hacen:

1) La confusión que se produce en los no preparados al puntuar con valores tan pronto negativos como positivos.

2) El tamaño excesivo de la unidad (sigma) elegida.

Hemos de anticipar el hecho de que estas críticas no se pueden dirigir contra el procedimiento en sí ni contra su eficiencia, sino más bien contra la sencillez de aplicación, contra los errores mecánicos y contra la facilidad interpretativa cuando se ofrecen a estudiosos con actitud negativista ante lo más matemático. Así ¿qué entenderíamos por $Z = -2,25$? Es posible que surjan espontáneas preguntas como éstas: ¿Es que el sujeto sabe menos que nada? ¿Es probable un escolar con \bar{Z} negativo?... Con estas preguntas se intentaría desvirtuar el procedimiento por aquellos para quienes el criterio de autoridad, apoyado en base endeble, es más valioso que la inferencia experimental. Estos interrogantes sólo pueden aparecer cuando el demandante olvida que por un proceso estadístico hemos conseguido que el valor medio sea nulo. La nulidad para ser conseguible exige o que todas las puntuaciones sean cero o que algunas sean menores de cero.

Es cierto que en la computación mecánica un cambio de signo influye con exceso y no es raro aparezca, pero hemos de reconocer la gran dificultad que encierra la eliminación por medio de pruebas de todo error mecánico. Es laudable el propósito, y si existe una vía que reduzca estos errores, debe ser utilizada.

No es difícil descubrir esta vía. Si para transformar en cero la media hemos restado a cada puntuación bruta el valor de la media, para convertir todos los Z en positivos será suficiente añadir a cada uno un valor constante de tal modo que la suma de esa constante con el menor valor posible de Z totalice cero. Dado que la probabilidad de aparición de un $|Z|$ mayor que 5 es menor de 29×10^{-8} , es decir, dado que de cada cien millones de escolares con pruebas indefinidas, aparecerían, probablemente, menos de 30 con una $|Z|$ mayor, puede tomarse como cons-

tante a añadir a cada Z hallada el número 5. Bajo la anterior probabilidad, se afirma ahora que las puntuaciones típicas oscilan entre 0 y 10 con media igual a 5.

El nuevo valor Z'_n sería igual a $Z'_n + 5$. En los casos de $Z'_a = -2,25$ $Z'_b = 1,84$ obtendríamos $Z'_a = -2,25 + 5 = 2,75$ $Z'_b = 1,84 + 5 = 6,84$

No encontramos suficientemente justificada la crítica de amplitud de la unidad como no sea apoyándose en criterios tradicionales o en las puntuaciones brutas de 0 a 100. ¿Acaso es más difícil comprender la diferencia entre dos escolares cuando las puntuaciones vienen dadas en la forma 2,45 y 6,82 que cuando se nos presentan como 24,5 y 68,2? Si los examinadores están enterados, las significaciones son equivalentes. Sólo es aceptable cuando se piensa manifestar al público, desconocedor del procedimiento, los resultados de los exámenes. Se evitaría así toda crítica respecto de la precisión de los juicios y de eliminación o aprobación por una fracción de punto.

De acuerdo con esta perspectiva social, y después de la escala T de Mac Call, se han variado los sistemas Z y Z' . Es suficiente multiplicar cada puntuación tipificada Z' por 10, con lo que convertimos la sigma en 10. La hipótesis en que se apoya respecto de la escala T, es: «la normalidad puede admitirse siempre que los instrumentos de medida sean fieles». Operamos ahora con una distribución supuesta normal con $\bar{X} = 50$ y $S = 10$. Si partiésemos de las Z iniciales, el nuevo valor Z'' se obtendría así:

$$Z'' = 50 + 10 Z$$

Si quisiéramos obtener directamente las puntuaciones Z'' sin averiguar las Z , hemos de recordar que

$$\frac{X' - \bar{X}'}{S'} = \frac{X - \bar{X}}{S}$$

Luego

$$X' = \left(\frac{S'}{S}\right) (X - \bar{X}) + \bar{X}' = \frac{(X - \bar{X}) 10}{S} + 50 = \frac{10}{S} X - \frac{10 \bar{X}}{S} + 50 = Z''$$

En esta fórmula $\frac{10}{S}$ es constante y $\frac{10 \bar{X}}{S}$ es también constante. Cabe construir una tabla similar a la tabla I que nos facilite el hallazgo de los valores Z'' .

Mas todo este procedimiento opera con otros supuestos:

- a) Identidad de valor ponderativo de todas las materias escolares.
- b) Las covarianzas son idénticas, y las correlaciones, iguales a uno.

La renuncia al primer supuesto nos llevaría, ora al cambio de las sigmas por hipótesis ponderativa, ora al pesaje de las puntuaciones mediante los coeficientes de regresión.

La aceptación del segundo, se traduce en reducción de algunas posibilidades de cada individuo al determinar su emplazamiento porcentual.

JOSÉ FERNÁNDEZ HUERTA

Profesor de la Universidad de Madrid

S U M M A R Y

Both in the field of School Organization and in that of Methodology we are confronted with the acute problem of the composition and unification of marks. In a summary way the author presents some defective procedures and studies the «standard scores method». He shows its evident superiority to other methods and he also offers some illustrative examples. The conditions and limitations of its applicability when we wish to compose strictly scientific scores are also studied. Finally Dr. Fernández Huerta offers a transformation of the «standard scores» which facilitates their understanding and lessens some mechanical errors.