

Análisis comparativo de formas verbales e icónicas en la medición del rendimiento

por
Antonio R. Bartolomé

INTRODUCCIÓN

El uso de elementos icónicos en ítems en diferentes tipos de pruebas y tests no es novedoso. Sin embargo, en las pruebas utilizadas habitualmente en los diferentes niveles educativos no es habitual el uso de figuras. En un análisis sobre 4 pruebas de rendimiento tipo test utilizadas en la asignatura «Estadística aplicada a las Ciencias de la Educación» menos del 5% de los ítems incluían elementos icónicos.

¿Cómo es que el uso masivo de mensajes icónicos en el proceso de aprendizaje no se corresponde con un uso similar en el momento de la evaluación? Apuntamos dos explicaciones:

- a) a nivel universitario la imagen tiene un papel subordinado a la comunicación verbal;
- b) un mayor uso de imágenes en las pruebas de rendimiento no alteraría sustancialmente los resultados académicos; esto junto con el mayor costo de realización haría poco rentable su uso.

En este trabajo se estudia esta segunda hipótesis.

ALGUNAS REFERENCIAS TEÓRICAS

La presentación de un ítem en forma verbal o icónica no es indiferente. La relación entre imagen y pensamiento ha sido abordada desde una perspectiva piagetiana (Sinclair, H. 1978), Psicométrica (Guilford, J.P. 1977) o desde una concepción de la inteligencia como «procesadora de la información» (Wagner, R.K. y Sternberg, R.J. 1984). Recientes trabajos en Psicología de Gringer y

Bandler (Patton M.Q. 1981) sugieren que personas diferentes han desarrollado diferentes sistemas (verbales, táctiles, sonoros y visuales) de representación e interpretación del mundo. M.Q. Patton concluye que no hay una modalidad de comunicación y experiencia única y efectiva para todos.

La imagen *no* siempre ayuda a resolver correctamente un ítem. Experiencias de Wohlwill y el trabajo de Winer y Kronberg (Denis, M. 1984) recogen algunos ejemplos en esta línea.

En nuestro caso, la introducción de elementos icónicos está referida a la codificación efectiva de información fundamental del ítem mediante códigos visuales. No se incluye el uso del comic como elemento motivador, siguiendo en ese sentido a Macdonald-Ross, M. (1978).

PLANTEAMIENTO DE LA INVESTIGACIÓN

El instrumento utilizado fue un test elaborado desde la perspectiva de una evaluación basada en criterios, concretamente siguiendo los planteamientos de Popham W.J. (1980, 1983). A partir de unas especificaciones se establecieron para cada una unos contenidos equivalentes: sobre este material se elaboraron ítems apareados, uno en forma verbal y otro en forma icónica. (*figura 1*).

Los ítems fueron asignados aleatoriamente-apareados a dos modelos (A y B) de la prueba. Cada modelo recogía ítems correspondientes a las distintas especificaciones, pero para una misma especificación, los ítems pertenecían a diferentes contenidos para evitar fenómenos de resonancia entre las formas icónica y verbal.

Los dos modelos se distribuyeron aleatoriamente según un esquema sistemático sobre la distribución espacial. La muestra era un grupo académico de 46 estudiantes de 2.º curso de Pedagogía. También se aplicó a este grupo una prueba clásica de problemas.

La muestra no tiene carácter de representatividad. Se trata de un estudio puntual destinado a ser replicado y mejorado en diferentes situaciones. Las técnicas inferenciales utilizadas en el análisis son tales que permiten utilizarlo como elemento metodológico en clase con los mismos alumnos.

ELABORACIÓN DE LA PRUEBA


El test pretendía medir un dominio preciso: la aplicación por parte de los estudiantes de técnicas inferenciales de estimación y contraste en base a las distribuciones Normal, t-Student y ji-cuadrado. La prueba se construyó según muestra el siguiente esquema:

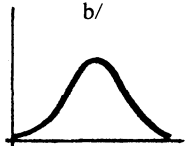
cuatro pruebas tipo test
utilizadas otros años
(análisis de contenido)


fig. 1

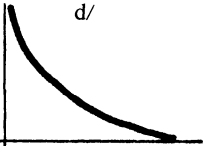
CONTENIDO

5. ¿Cuál de entre los siguientes modelos de distribuciones de probabilidad sigue la distribución muestral de una proporción?

a/


b/


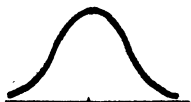
c/


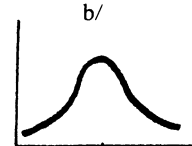
d/



5. ¿Cuál de entre los siguientes modelos de distribuciones de probabilidad sigue la distribución muestral de una proporción?

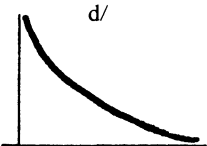
a/ Ley Normal
b/ Ley Normal asimétrica
c/ distribución t con 1 grado de libertad
d/ distribución chi-cuadrado

13. ¿Cuál de entre los siguientes modelos de distribuciones de probabilidad sigue la distribución muestral de una media?

a/


b/


c/


d/


13. ¿Cuál de entre los siguientes modelos de distribución de probabilidad sigue la distribución muestral de una media?

a/ Ley Normal asimétrica
b/ t de Student con x grados de libertad
c/ chi-cuadrado
d/ F de Snedecor

I.

1. Con referencia a las *distribuciones muestrales* correspondientes a

- . proporción en que un carácter se encuentra presente
- . distribución de sujetos entre varias categorías
- . media de una variable cuantitativa

identificar:

- . modelo de probabilidad que sigue

listado de 16 destrezas

(reagrupación; comparación con objetivos de la asignatura)

listado de 13 destrezas básicas

(encuesta valorativa entre 4 profesores de la asignatura)

ESPECIFICACIONES

(redacción a partir de especificaciones y temario)

(evaluación y corrección por jurado de profesores del Departamento)

Contenidos alternativos de dificultad presuntamente equivalente

(elaboración de ítems por parejas, con formas icónicas y verbales)

ÍTEMS

(asignación aleatoria por parejas)

MODELOS A y B

La prueba fue fiabilizada por un procedimiento similar al sugerido por Popham W.J. (1983, pg. 206). El criterio de decisión se estableció en un 70% de aciertos («experto»). Se obtuvo un coeficiente estimado entre .551 y .601 (alfa = .05). Al existir dos modelos (A y B) también se compararon ambos. Dado el carácter del test se prefirió una prueba no-paramétrica (U de Mann-Whitney), obteniendo zeta = .76 (grado de signif. aprox. .45).

La validez descriptiva se trabajó en el proceso de elaboración de las especificaciones; este tipo de validez es el más importante para la medición basada en criterios, según Mehrems W.D. (1973) y fue especialmente cuidada. Desde un punto de vista funcional se estudió la correlación entre la calificación («experto», «no-experto») en el test y en una prueba clásica de problemas. Se obtuvo un valor Phi = .323 (significativo al nivel .05). Estudiando las fuentes de error debidas al diferente carácter de ambas pruebas se señalan:

- a) existencia de cálculos en la prueba de problemas
- b) posibilidad de aciertos al azar en el test.

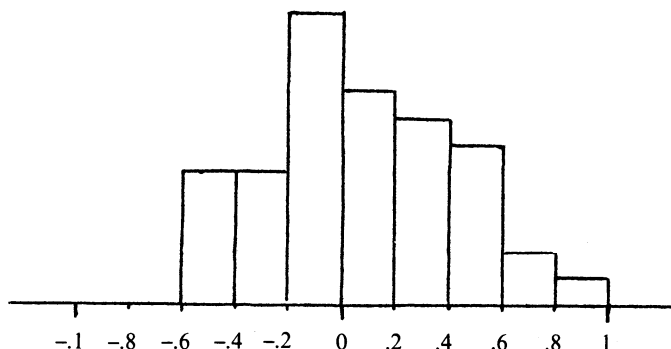
Notemos que en ambos casos se tiende a incrementar el error tipo II, y a reducir el valor numérico de Phi.

No nos detenemos más en el análisis de los ítems, remitiendo a lo que indica Popham W.J (1983, pg. 140) sobre la necesidad de una homogeneidad derivativa y no una funcional. Constatemos de cualquier modo la escasez de mecanismos eficaces y seguros en este terreno, al menos por comparación con los tests basados en normas (Haertel, E. 1985).

ANÁLISIS DE RESULTADOS

Inicialmente se calcularon los coeficientes Phi entre los aciertos en una u otra forma (icónica-verbal) para cada individuo de la muestra. Sólo en 3 de los 46 se encontraron coeficientes significativos (*figura 2*). A la vista de estos resul-

Figura 2



(coeficientes PHI para cada individuo de la muestra entre la calificación obtenida en ítems icónicos y la obtenida en verbales)

tados podría pensarse que existen diferencias entre el modo como responden los alumnos en uno u otro caso. Sin embargo conviene hacer dos matizaciones.

En primer lugar, siguiendo a Popham, en un test referido a criterios el peso se desplaza del ítem al conjunto de la prueba. Así no interesa tanto constatar si cuando un alumno acierta un ítem en forma icónica, acierta también otro similar presentado en forma verbal, sino si los alumnos que serían considerados expertos en base al conjunto de los ítems verbales, también lo serían en base a los ítems icónicos. Aplicando un coeficiente Phi para el conjunto de la muestra según lo indicado obtenemos .5167 (significativo al nivel .05). En el 77% de los casos un alumno obtenía la misma calificación basándose en los ítems icónicos o en los verbales.

En segundo lugar recordemos que cuando apareamos los ítems correspondientes a una misma destreza, para cada alumno, no correspondían únicamente a diferentes formas (verbal o icónica) sino también a diferentes contenidos (supuestamente equivalentes). Por ello se procedió a analizar los índices de dificultad de los ítems según destrezas, contenidos y formas. A partir de estos datos se contrastaron las proporciones de aciertos entre formas (dentro de un mismo contenido) y entre contenidos (dentro de una misma forma). Se encontraron 4 diferencias significativas en el primer caso, y 9 en el segundo (nivel .05). Las diferencias entre contenidos se acentuaron cuando los ítems los presentaban en forma verbal (7/10 diferencias significativas).

Los ítems en los que se encontraron diferencias significativas por algunos de los conceptos anteriores fueron analizados, obteniéndose algunos resultados interesantes. Alguno se ha incluido en las conclusiones.

El índice de aciertos general fue de .61 para la forma icónica y .63 para la verbal. La distribución de aciertos entre diferentes destrezas, tanto para los contenidos I como los II, no mostraron diferencias significativas entre las formas (grados de significación de 0.50 y 0.80).

Conclusiones

Para el conjunto de la prueba, las diferencias encontradas no afectan sustancialmente los resultados, si bien invitan a seguir profundizando en la causa de las diferencias.

En ítems concretos se encuentran diferencias muy significativas que sugieren que las especificaciones normalmente deberían incluir indicaciones sobre la conveniencia de utilizar imágenes o no en el ítem. Por otro lado, la misma presentación de los textos o datos actuó como una «imagen» alterando la percepción del ítem.

La falta de homogeneidad entre contenidos supuestamente equivalentes lleva a destacar la eficacia del azar para elaborar pruebas equivalentes (por ejemplo, a partir de un banco de ítems) frente al clásico sistema de elaborar similares. Parece que la forma icónica suaviza las diferencias entre contenidos.

En esta línea podría estudiarse si las formas verbales discriminan más, y si esa discriminación es relevante.

En próximos años, esta prueba se aplicará a alumnos que utilizarán intensivamente el vídeo, lo que permitirá estudiar los cambios que se produzcan en los resultados.

REFERENCIAS BIBLIOGRÁFICAS

- DENIS, M.: (1984) *Las imágenes mentales*. Siglo XXI, Madrid.
- GUILFORD, J.P.: (1977) *Way beyond the I.Q.* The Creative Education Foundation Inc. Buffalo.
- HAERTEL, E.: (1985) «Construct Validity and Criterion-Referenced Testing», en *Review of Educational Research*, vol. 55, N.º 1, pp. 23-46.
- MACDONALD-ROSS, M.: (1978) «Graphics in texts», en *Review of Research in Education* 5, pp. 49-85. F.E. Peacock Publishers, Itasca (Illinois).
- MEHRENS, W.A.: (1973) *Measurement and Evaluation in Education and Psychology*. Holt, Rinehart and Winston, New York.
- PATTON, M.Q.: (1981) *Creative Evaluation*. Sage Publ., London.
- POPHAM, W.J.: (1980) *Problemas y técnicas de la evaluación Educativa*. Anaya, Madrid.
- POPHAM, W.J.: (1983) *Evaluación basada en criterios*. Edit. Magisterio Español, Madrid.
- SINCLAIR DE ZWART, H.: (1978) *Adquisición del lenguaje y desarrollo de la mente*. Oikos Tau. Barcelona.
- WAGNER, R.K. y STERNBERG, R.J.: (1984) «Alternative Conceptions of Intelligence and their implications for Education» en *Review of Educational Research*, vol. 54, n.º 2. Washington.