

Mejora de las puntuaciones de los tests mediante el entrenamiento

Julia Martínez-Cardeñoso, José Muñiz y Eduardo-García Cueto
Universidad de Oviedo

Las pruebas compuestas por ítems de elección múltiple tienen una serie de características específicas que permiten obtener mejores resultados cuando, además del conocimiento en la materia, se dispone de determinadas habilidades y se utilizan determinadas estrategias que se adquieren con la práctica y la experiencia. En este trabajo se analiza el efecto de dos programas de entrenamiento sobre las puntuaciones de las personas en los tests. El primer entrenamiento fue diseñado para entrenar a las personas en la realización de tests destinados a la evaluación de conocimientos académicos. Es un entrenamiento basado fundamentalmente en la enseñanza de los conocimientos a evaluar. Se aplicó a una muestra de 152 alumnos de la Universidad de Oviedo, de los cuales 100 constituyeron el grupo experimental y 52 el grupo control. Los resultados muestran que el programa de entrenamiento ha resultado efectivo en la mejora de las puntuaciones de las personas entrenadas. El segundo entrenamiento utilizado consiste únicamente en el aprendizaje de estrategias de resolución de tests, con independencia de los contenidos. Dicho entrenamiento fue aplicado a una muestra de 120 trabajadores de la administración pública. Los resultados muestran con claridad que este segundo tipo de entrenamiento centrado en las estrategias también produce mejoras significativas en las puntuaciones de las personas en el test, observándose una mayor ganancia en aquellas personas con mayor nivel de estudios.

Coaching effects on test scores. The effect of two different coaching programs on test scores was analyzed. The first program was designed to coach individuals who have to pass academic tests, in which academic knowledge was the variable measured. In this program the training was focused on the academic contents to be assessed. A sample of 152 university students was used, 100 as experimental group, and 52 as a control group. Results show that the coaching is effective, significantly increasing test scores. The second program is focused in the strategies and test wisdom used by the subjects in order to answer the items. A sample of 120 workers from the public administration was used. The program effects are clear, increasing significantly individuals test scores. Finally, the implications, limitations, and possibilities of both coaching programs are discussed.

En la actualidad se dispone de muy diversos programas para entrenar a las personas en la realización de los tests (Anastasi, 1981; Deaton, Halpin y Alford, 1987; Messick y Jungesblut, 1981; Millman, Bishop y Ebel, 1965; Powers, 1993; Powers y Leung, 1995). No puede hablarse con rigor de programas óptimos, pues la mejora de las puntuaciones de los tests debido a los entrenamientos varía en función a las estrategias empleadas, las cuales serán más efectivas dependiendo del tipo y naturaleza de los ítems (Johnson y Wallace, 1989; Powers, 1986), de las características y

formato de los tests (Millman, Bishop y Ebel, 1965; Powers y Leung, 1995), del contenido del test (Bond, 1989; Powers y Swinton, 1984; Schiano y Kahleifeh, 1988) y de la naturaleza de los tests (Anastasi, 1981; Bond, 1989; Slack y Porter, 1980). Otros aspectos considerados por algunos autores como influyentes en la efectividad de los entrenamientos son el tiempo de duración de los entrenamientos (Coffman y Parry, 1967; Messick y Jungesblut, 1981; Pallone, 1961) y las diferencias individuales (Johnson, Asbury, Wallace, Robinson y Vaughn, 1985; Messick y Jungesblut, 1981; Millman, Bishop y Ebel, 1965; Scruggs, Bennion y Lifson, 1985 a y b)

El objetivo de esta investigación es analizar el efecto de dos tipos de programas de entrenamiento sobre las puntuaciones de las personas en los tests. El primer programa va encaminado a mejorar el rendimiento académico y se basa principalmente en la enseñanza de los contenidos de la materia a evaluar. Por el contrario, el

segundo programa se centra en el entrenamiento de estrategias para resolver con éxito ítems de elección múltiple.

Estudio 1

El objetivo de este primer estudio es determinar el efecto del entrenamiento basado en la instrucción sobre las puntuaciones de las personas en el test.

Método

Participantes

Se utilizó una muestra de 152 alumnos voluntarios de segundo curso de la Facultad de Psicología de la Universidad de Oviedo.

Entrenamiento

El entrenamiento constaba de una combinación de estrategias de resolución de tests, revisión de la materia, clases docentes, horas de tutoría y horas de estudio de los alumnos. Desde el punto de vista de Messick y Jungeblut (1981), se trataría de un tipo de preparación cuyo objetivo es la instrucción intensiva enfocada a desarrollar el constructo y conocimientos.

A las personas que participaron en la investigación se les aplicó una prueba de evaluación de conocimientos psicométricos (pretest), utilizando como postest el examen oficial de la asignatura. Entre ambas pruebas se llevó a cabo el entrenamiento. El entrenamiento constaba de siete elementos: estrategias de reducción de ansiedad, familiarización con los tests y situación de examen, feedback, enseñanza y práctica de estrategias de distribución eficaz del tiempo del examen, estrategias para evitar errores, estrategias para responder las preguntas por azar y estrategias de razonamiento deductivo. Estos elementos fueron explicados y puestos en práctica a lo largo de cuatro sesiones de entrenamiento de una hora, finalizadas las sesiones se les hizo el examen parcial de la asignatura (postest).

Resultados

Al aplicar el programa de entrenamiento ocurrió que de los 152 alumnos que participaron en la investigación, 100 realizaron alguna sesión de entrenamiento además del pretest y postest, 17 realizaron las pruebas de pretest y postest pero ninguna sesión de entrenamiento y 35 realizaron sólo el postest; en consecuencia, para examinar de forma general el efecto del entrenamiento, se calcularon las puntuaciones medias en el pretest y postest del grupo formado por los 100 alumnos que habían participado en alguna sesión de entrenamiento, obteniendo los resultados de la Tabla 1.

En la tabla 1 se observa una marcada diferencia entre las medias obtenidas por los alumnos en los dos exámenes. Para deter-

minar si estas diferencias son significativas se realizó un análisis de varianza de medidas repetidas, obteniendo un valor de $F=686,42$ $p<0,001$. En consecuencia, globalmente el entrenamiento realizado resulta efectivo.

Un aspecto de sumo interés es el papel que juega la duración del entrenamiento, siendo por ello una de las variables más estudiadas por los investigadores. Debido a que la asistencia de los alumnos a las sesiones de entrenamiento era voluntaria, se obtuvo la siguiente distribución: a) 35 alumnos no realizaron el pretest sólo el postest, b) 17 alumnos realizaron solamente el pretest y postest, es decir, ninguna sesión de entrenamiento, c) 49 alumnos realizaron el pretest y postest y una sesión de entrenamiento d) 23 alumnos realizaron el pretest, postest y dos sesiones de entrenamiento, e) 13 alumnos realizaron el pretest, postest y tres sesiones de entrenamiento y f) 15 alumnos realizaron el pretest, postest y cuatro sesiones de entrenamiento.

En consecuencia, se pueden analizar las diferencias entre las puntuaciones de los participantes en el pretest y postest considerando el número de sesiones en las que participaron, si bien hay que tener en cuenta que los resultados de este estudio deben ser considerados con precaución debido a dos aspectos no controlados como son que la distribución de los alumnos en los grupos no es estrictamente aleatoria, y que el número de alumnos por grupo es bajo.

Las puntuaciones medias de cada uno de los grupos definidos anteriormente, se detallan en el gráfico 1.

Como se puede observar en el gráfico 1, en el pretest, la puntuación media más elevada es la de los alumnos que realizaron el pretest y postest pero ninguna sesión de entrenamiento (grupo 0) con una media de 13,80, seguida de la media del grupo de alumnos que recibió una sesión de entrenamiento (grupo 1) cuya media es 13,79. En el caso del postest se observan unas puntuaciones medias muy superiores a las obtenidas en los grupos del pretest, pero el mayor aumento se observa, en este caso, en el grupo que ha asistido a las cuatro sesiones de entrenamiento (grupo 4) con una puntuación media de 35.

El análisis de las diferencias entre el pretest y postest, considerando el efecto de las sesiones de entrenamiento, se realizó mediante un análisis de varianza de dos factores con medidas repetidas en un solo factor, siendo la variable intersujeto el número de sesiones a las que asistieron los alumnos y cuyos valores oscilan de cero (grupo control) a cuatro (número máximo de sesiones de entrenamiento a las que podían asistir los alumnos). En dicho análisis de varianza se obtuvo un valor de $F(4,112) = 0,20$ y $p = 0,936$ en la variación intergrupos, es decir, no existen diferencias entre los grupos. En la variable intragrupo se obtuvo un valor de $F(1,112) = 763,37$ y $p < 0,001$, es decir, existen diferencias significativas entre el pretest y postest. La interacción resultó ser también significativa, con un valor de $F(4,112) = 3,88$ $p < 0,005$.

Discusión y conclusiones

Aunque con cautela, debido a las limitaciones previamente señaladas, los resultados obtenidos indican que el entrenamiento es efectivo. Sin llegar a niveles de significación estadística, también se observa una conexión entre el número de sesiones de entrenamiento y la puntuación: a mayor número de sesiones corresponde una puntuación más elevada. Estos resultados concuerdan con los obtenidos por Messick y Jungeblut (1981), Dyer (1953) o Roberts y Oppenheim (1966).

Tabla 1

Puntuación media en el pretest y postest de los alumnos que recibieron algún tipo de entrenamiento

	N	Media	Desviación Típica
Pretest	100	13,07	5,97
Postest	100	31,97	7,45

Otro aspecto a destacar del entrenamiento realizado es que las personas que puntuaron más bajo en el pretest son las que, posteriormente, asistieron al mayor número de sesiones de entrenamiento y obtuvieron puntuaciones más altas en el examen posttest, lo que corrobora la idea de Anastasi (1981) y Bond (1989) según la cual los estudiantes que utilizan estrategias poco efectivas en la realización del test parecen beneficiarse más.

Estudio 2

En esta investigación se trata de averiguar los efectos de un programa de entrenamiento centrado en diversas estrategias, pero no en los contenidos, sobre las puntuaciones en un test. Al contrario de lo que ocurría en el primer estudio, en el que el programa de entrenamiento se centraba en los contenidos a evaluar, aquí se trata de un programa que prescinde de los contenidos y se centra en estrategias varias encaminadas a mejorar el rendimiento de los sujetos en los ítems de elección múltiple. Al lado de la efectividad del programa, también se estudiarán aspectos tales como la incidencia del nivel de estudios y la edad de las personas entrenadas.

Método

Participantes

En este programa de entrenamiento participaron 120 personas aspirantes trabajadores de la administración pública asturiana.

Entrenamiento

El programa de entrenamiento aplicado a las personas del curso se desarrolló en 21 horas, 3 horas cada día. Durante los siete días que duró se les entrenó en siete técnicas fundamentales: estrategias de reducción de ansiedad, familiarización con los tests y situación de examen, feedback, enseñanza y práctica de estrategias de distribución del tiempo del examen eficazmente, aprendizaje de las estrategias para evitar errores, aprendizaje de estrategias para responder las preguntas al azar y uso de estrategias de razonamiento deductivo.

Estas estrategias fueron explicadas y puestas en práctica a lo largo de las siete sesiones de entrenamiento. Así mismo, y para que las personas practicaran las estrategias que iban aprendiendo, se les facilitó una batería de tests psicotécnicos para practicar, la cual era corregida y analizada por los entrenadores, analizando cuales eran las estrategias más indicadas en cada caso. Nótese que aquí todo el entrenamiento era independiente de los contenidos de las pruebas, no entrenando directamente sobre ellos. Para evaluar el efecto del entrenamiento diseñado se utilizó como pretest y posttest una misma prueba psicotécnica de 100 preguntas. Este examen fue aplicado el primer y último día de entrenamiento, con un tiempo máximo para realizarlo de 75 minutos.

Resultados

Las puntuaciones medias de las personas en el pretest y posttest se pueden ver en la tabla 2.

La tabla 2 refleja la diferencia entre la puntuación media del pretest y la del posttest (18,87 y 26,17, respectivamente). Esta diferencia resultó ser estadísticamente significativa [$F(1,119) = 99,40$; $p < 0,001$] cuando se aplicó un análisis de varianza de medidas repetidas.

En el estudio de la relación entre el nivel de estudios de las personas participantes y el entrenamiento se ha podido comprobar que la mayoría de los participantes en el entrenamiento habían realizado estudios primarios (81) o bachiller (29), tan sólo una persona había realizado estudios superiores, otras tres personas realizaron otros estudios y seis personas no especificaron los estudios realizados. La aplicación del entrenamiento a las personas partici-

	N	Media	D.T.
Pretest	120	18,87	7,19
Posttest	120	26,17	8,75

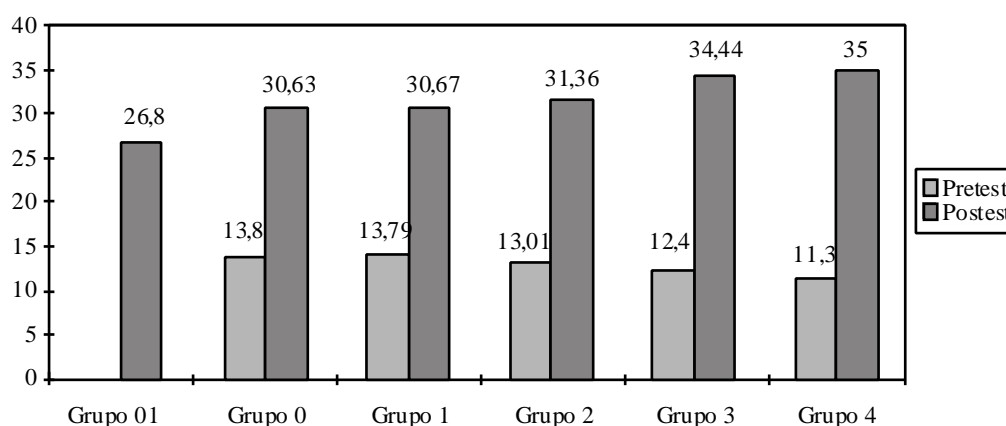


Gráfico 1. Puntuación media en el pretest y posttest de los distintos grupos:

Grupo 01 - Alumnos que realizaron sólo el posttest

Grupo 0 - Alumnos que realizaron el pretest y posttest, pero ninguna sesión de entrenamiento

Grupo 1 - Alumnos que realizaron pretest, posttest y asistieron a una sesión de entrenamiento

Grupo 2 - Alumnos que realizaron pretest, posttest y asistieron a dos sesiones de entrenamiento

Grupo 3 - Alumnos que realizaron pretest, posttest y asistieron a tres sesiones de entrenamiento

Grupo 4 - Alumnos que realizaron pretest, posttest y asistieron a las cuatro sesiones de entrenamiento

pantes en la investigación produjo un aumento de la puntuación en el postest, con relación al pretest, en todos los grupos determinados por el nivel de estudios como puede verse en la tabla 3.

Para determinar el efecto del nivel de estudios, además del efecto del entrenamiento, se realizó un análisis de varianza de dos factores con medidas repetidas en uno de los factores. La variable intersujeto será el nivel de estudios, que en este caso tendrá dos niveles uno para estudios primarios y otro para estudios de bachiller debido a que los niveles de estudios superiores y otros estudios se eliminaron por haber pocas personas en cada grupo. Los resultados obtenidos en dicho análisis de varianza fueron: 1) En la variable intersujetos se obtuvo un valor de $F(1,109) = 24,48$ $p < 0,001$. 2) En la variable intrasujetos $F(1,109) = 24,48$ $p < 0,001$ y 3) en la interacción $F(1,109) = 24,48$ $p < 0,001$. A la vista de los resultados se observa que las diferencias entre los niveles de estudio resultan estadísticamente significativas.

Las edades de las personas que realizaron el entrenamiento estaban comprendidas entre 29 y 60, tal como se indica en el gráfico 2.

Debido al reducido número de personas en algunos de los grupos, para realizar los análisis de varianza se distribuyeron en tres grupos (tabla 4): a) Personas con edades menores de 40 años b) Personas con edades comprendidas entre 40 y 49 años c) Personas mayores de 49 años.

Al comparar los grupos mediante la prueba de Scheffe se obtuvieron diferencias significativas en el postest entre los grupos de las personas menores de cuarenta años y las personas de edad comprendidas entre 40 y 49 años $p < 0,001$; también se encontraron diferencias entre las personas menores de cuarenta años con las personas mayores de cincuenta años $p < 0,001$.

Discusión y conclusiones

Para interpretar cabalmente los resultados obtenidos en este estudio hay que tener presentes dos aspectos que atenúan la solidez de las conclusiones. En primer lugar, no ha sido posible utilizar grupo control y en segundo, la dificultad del test utilizado como pretest y postest resultó ser muy elevada para la muestra de personas participantes. Aún así, el aumento significativo en la puntuación indica que el entrenamiento en estrategias realizado resultó ser efectivo, por lo que es posible que con un examen de dificultad más ajustada a la muestra empleada se detectasen diferencias incluso mayores. Estos resultados concuerdan con los obtenidos por Callenbach (1973), Lifson, Scruggs y Bennion (1984) y Samson (1985). Dos variables que en principio parecen influir en la efectividad del entrenamiento son la edad (a más edad menor aumento en la puntuación) y el nivel de estudios (a mayor nivel de estudios mayor aumento en la puntuación). Estos resultados corroboran los planteamientos de Anastasi (1981), Messick y Jungeblut (1981) y Bond (1989), según los cuales el entrenamiento puede tener un efecto diferencial en los distintos estudiantes.

Por otra parte, se puede concluir, al igual que en las investigaciones de Jones (1986 a y b), que las personas de más baja puntuación en el pretest tienden a beneficiarse menos del entrenamiento. Este aspecto, que contrasta con los resultados del estudio 1, puede tener explicación si consideramos, al igual que Evans y Pike (1973), Pike y Evans (1972), Slack y Porter (1980), Johnson y Wallace (1989), Linn (1990) y Powers (1993), que la naturaleza de los tests y el tipo de entrenamiento pueden tener un papel importante en la efectividad del mismo.

Tabla 3
Variación de la puntuación media entre el pretest y postest con relación al nivel de estudios

Nivel de estudios	N	Pretest		Postest	
		Media	D. T.	Media	D. T.
E. Primarios	81	17,95	7,38	23,14	7,26
Bachiller	29	21,03	5,88	33,96	8,21
E. Superiores	1	14,00		28,00	
Otros	3	19,66	3,78	28,00	4,35

Tabla 4
Puntuaciones medias y desviación típica de los grupos de edades

	N	Pretest		Postest		Δ Media
		Media	D.T.	Media	D.T.	
< 40	37	21,86	7,79	31,35	7,18	9,49
40-49	57	17,56	6,60	24,12	8,90	6,56
>=50	19	17,57	6,56	22,57	7,47	5,00

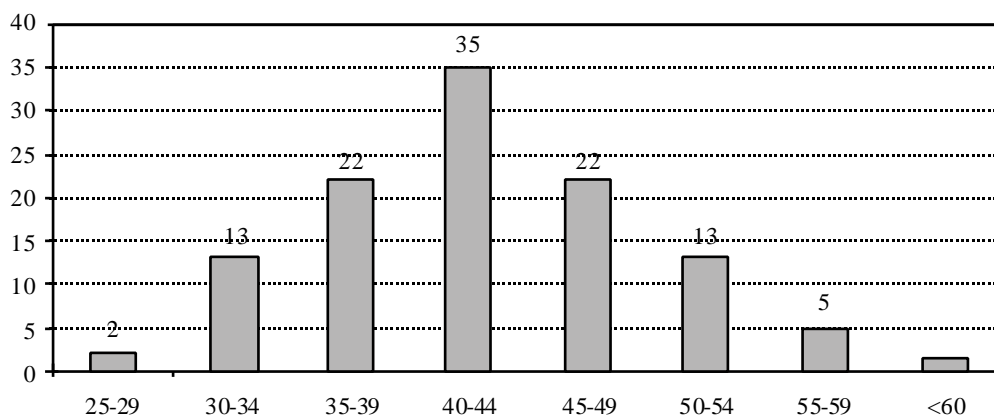


Gráfico 2. Distribución de las personas por edades

Referencias

- Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist*, 36(10), 1086-1093.
- Bond, L. (1989). The effects of special preparation on measures of scholastic ability. En: R. Linn. *Educational Measurement*. New York: MacMillan Publishing Company.
- Callengach, D. (1973). The effects of instruction and practice in content-independent test-taking techniques upon the standardized reading test scores of selected second-grade students. *Journal of Educational Measurement*, 10, 25-30.
- Coffman, W. E. y Parry, M. E. (1967). Effects of an accelerated reading course on SAT-V scores. *Personnel and Guidance Journal*, 46, 292-296.
- Deaton, W. L., Halpin, G. y Alford, T. (1987). Coaching effects on California achievement test scores in elementary grades. *Journal of Educational Research*, 80(3), 149-155.
- Dyer, H. S. (1953). Does coaching Help? *College Board Review*, 19, 331-335.
- Evans, F. R. Y Pike, L. W. (1973). The effects of instruction for three mathematics item formats. *Journal of Educational Measurement*, 10, 257-272.
- Johnson, S. T., Asbury, C. A., Wallace, M. B., Robinson, S. y Vaughn, J. (1985). The effectiveness of a program to increase Scholastic Aptitude Test scores of black students in three cities. *Paper presented at the Annual Meeting of the National Council on Measurement in Education*. Chicago.
- Johnson, S. T. Y Wallace, M. B. (1989). Characteristics of SAT quantitative items showing improvement after coaching among black students from low-income families. An exploratory study. *Journal of Educational Measurement*, 26(2), 133-145.
- Jones, R. F. (1986a). A comparison of the predictive validity of the MCAT for coached and uncoached students. *Journal of Medical Education*, 61(4), 335-338.
- Jones, R. F. (1986b). The effect of commercial coaching courses on performance on the MCAT. *Journal of Medical Education*, 61(4), 273-284.
- Lifson, S., Scruggs, T. y Bennion, K. (1984). Passage independence in reading achievement tests: a follow-up. *Perceptual and Motor Skills*, 58, 945-946.
- Linn, R. L. (1990). Admissions testing: Recommended uses, validity, differential prediction, and coaching. *Applied Measurement in Education*, 3(4), 297-318.
- Messick, S. y Jungeblut, A. (1981). Time and method in coaching for the SAT. *Psychological Bulletin*, 89(2), 191-216.
- Millman, J., Bishop, Ch. Y Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement*, 25, 707-726.
- Pallone, N. J. (1961). Effects of short-term and long-term developmental reading courses upon SAT verbal scores. *Personnel and Guidance Journal*, 39, 654-657.
- Pike, L. W. Y Evans, F. R. (1972). The effects of special instruction for three kinds of mathematics aptitude items. *College Board Research Development*. Rep. N° 71-72.
- Powers, D. E. (1986). Relations of test item characteristics to test preparation/test practice effects: a quantitative summary. *Psychological Bulletin*, 100(1), 67-77.
- Powers, D. E. (1993). Coaching for the SAT: Summary of the summaries and an update. *Educational Measurement: Issues and Practice*, 12, 24-30.
- Powers, D. E. y Leung, S. W. (1995). Answering the new SAT reading comprehension questions without the passages. *Journal of Educational Measurement*, 32(2), 105-129.
- Powers, D. E. y Swinton, S. S. (1984). Effects of self-study for coachable test item types. *Journal of Educational Psychology*, 76, 266-278.
- Roberts, S. O. Y Oppenheim, D. B. (1966). *The effects of special instruction upon test performance of high school students in tennessee* (CB RDR 66-7, N° 1 and ETS RB 66-36). Princeton, N. J.: Educational Testing Service.
- Samson, G. (1985). Effects of training in test-taking skills on achievement test performance: A quantitative Synthesis. *Journal of Educational Research*, 78(5), 261-266.
- Scruggs, T., Bennion, K. Y Lifson, S. (1985a). An analysis of children's strategy on reading achievement tests. *The Elementary School Journal*, 85(4), 479-484.
- Scruggs, T. Bennion, K. Y Lifson, S. (1985b). Learning disabled students' spontaneous use of test-taking skills on reading achievement tests. *Learning Disability Quarterly*, 8, 205-210.
- Slack, W. V. y Porter, D. (1980). The scholastic aptitude test: A critical appraisal. *Harvard Educational Review*, 50, 154-175.
- Schiano, D. Y Kahleifeh, B. (1988). *Spatial aptitude as expertise; a training study*. Paper presented at the Annual convention of the American Psychological Association, Atlanta, August 12-16.