La incertidumbre en la medida de la inteligencia

A. Martorell Cafranga^a y J. L. Ayuso Mateos^b

^a Fundación Carmen Pardo-Valcarce. ^b Departamento de Psiquiatría. Hospital Universitario de La Princesa. Universidad Autónoma de Madrid

Uncertainty on the measurement of intelligence

Resumen

En los últimos años se ha puesto de manifiesto que existe un patrón de incremento del cociente intelectual (CI) de 3 puntos por década en diversas poblaciones estudiadas. Esta observación se ha denominado efecto Flynn. Este artículo lleva a cabo una revisión bibliográfica de los datos aportados a propósito de este hallazgo y sus implicaciones en la evaluación de la inteligencia. El incremento antes enunciado plantea serios problemas, tanto metodológicos como teóricos, para el uso de los tests. En cuanto a los problemas metodológicos se analizan las consecuencias que el efecto Flynn tiene en la capacidad de los tests para dar una medida fiable de inteligencia, así como en lo referente a los estudios epidemiológicos que manejan datos del CI de poblaciones, en especial aquellos en los que se estudia la prevalencia del retraso mental. En lo referente a los problemas teóricos que el efecto Flynn pone de manifiesto, se aborda la cuestión de si los incrementos presentados por Flynn son incrementos reales de inteligencia o si se trata de algún efecto que distorsiona la medida. Se revisan las hipótesis que intentan dar cuenta de estos incrementos, en particular aquellas que apuntan a que existe un verdadero incremento generacional en lo que se refiere a la inteligencia, así como las que intentan explicar el efecto Flynn atribuyendo los incrementos a variables que distorsionan la medida, es decir, variables que incrementan los resultados de los tests pero que no incrementan la inteligencia de las poblaciones.

Palabras clave: Inteligencia. Neuropsicología. Evaluación de la inteligencia.

Summary

In recent years, there has been considerable evidence on the phenomenon of intelligence quotient (IQ) gains over time with a gain rate of approximately three IQ points per decade. This phenomenon has been called the Flynn effect. This review article presents the evidence that supports this effect and discusses its implications for the measurement of intelligence. The above mentioned IQ gains over time obviously present serious methodological and theoretical problems in the use of intelligence tests. We review the methodological consequences which the Flynn effect presents in the reliability of the measurement of IQ as well as the methodological consequences which this effect has for epidemiological studies, especially those which focus on the study of the prevalence of mental retardation. Mention is made of the hypothesis that tries to explain these IQ gains as real gains in intelligence, as well as those that propose that these gains are due to other causes rather to real population IQ gains.

Key words: Intelligence. Neuropsychology. Assessment of intelligence.

EL EFECTO FLYNN

Desde aproximadamente 1918 la media de las puntuaciones de los tests de inteligencia se ha visto incrementada consistente y significativamente¹. Una perspectiva amplia de este fenómeno ofrece una media de incremento de 3 puntos del coeficiente intelectual (CI)

por década, duplicándose casi este incremento en ciertas medidas especializadas. Países como Holanda, donde cada año se toman medidas de razonamiento abstracto como parte de las pruebas militares, hacen fácil la observación y medición de esta media de incremento.

Este fenómeno fue bautizado en 1994 por Herrnstein y Murray en su polémico libro *The bell curve* con el nombre de efecto Flynn². James R. Flynn fue el primero en enunciar este incremento cuando, al estudiar los resultados de las pruebas de Weschler para su polémica con el genetista Jensen, observó aumentos significativos en las puntuaciones de los tests de inteligencia a lo largo del tiempo. Investigaciones anteriores habían obviado este fenómeno por la tipificación de las puntuaciones directas en medidas de CI con su media estandarizada de 100.

Correspondencia:

José Luis Ayuso Mateos
Servicio de Psiquiatría
Hospital Universitario de La Princesa
Diego de León, 62
28006 Madrid
Correo electrónico: joseluis.ayuso@uam.es

Cuando Flynn retomó las puntuaciones directas y las comparó intergeneracionalmente observó asombrado un patrón constante de incremento en cualquier test de inteligencia o población estudiadas. Desde entonces se han sucedido las publicaciones de este autor, documentando siempre rigurosa y minuciosamente estos incrementos y proponiendo numerosos debates para conseguir una explicación aceptable del efecto Flynn.

Hasta ahora han sido 20 los países estudiados (Holanda, Bélgica, Francia, Noruega, Suecia, Dinamarca, Alemania, Austria, Suiza, Escocia, Irlanda del Norte, Inglaterra, Canadá, Estados Unidos, Australia, Nueva Zelanda, Israel, Brasil, Japón y China³), encontrando en todos ellos estos incrementos masivos del CI a lo largo del tiempo. Pero aunque siempre se han encontrado diferencias en las medias de incremento entre los distintos países, las diferencias más interesantes son las encontradas entre los distintos tipos de tests de inteligencia. Si analizamos el patrón de incremento de los últimos 60 años, las ganancias de CI han sido significativamente mayores en las pruebas que supuestamente son las medidas más puras de inteligencia, los tests de inteligencia fluida o medidas del factor g, que además se han considerado como medidas libres de sesgos culturales. Son pruebas de inteligencia fluida aquellas que miden la capacidad mental pura de resolución de problemas en un momento dado, independientemente de los conocimientos adquiridos. En el otro lado están las pruebas de inteligencia cristalizada que sí evalúan los conocimientos adquiridos, que presumiblemente variarán en función de una inteligencia más pura subyacente⁴. Las pruebas más reconocidas como medida de inteligencia cristalizada son las Escalas de Weschler, y como medida de inteligencia fluida las Matrices progresivas de Raven. Pues bien, es precisamente en estas últimas (pruebas de inteligencia fluida) donde se han hallado las mayores ganancias de CI; ganancias de aproximadamente 20 puntos por generación (30 años)^{1,5 (1)}. Las Escalas de Weschler muestran ganancias de entre 9 y 20 puntos por generación, mostrando la subescala verbal un aumento de 9 puntos como media. Entre los 11 países que ofrecen comparación entre ellos no hay uno solo en el que las ganancias de los tests «tipo Raven» no sean al menos dos veces mayores que las ganancias de las escalas «cristalizadas», encontrando ratios de 2:1 o 3:1⁵. Más aún, los subtests de aritmética, información y vocabulario no muestran incrementos entre generaciones^{5,7}. Es decir, las subescalas y los subtests que miden inteligencia cristalizada dentro de las pruebas de Weschler muestran siempre un incremento significativamente menor que aquellos que se correlacionan con medidas de inteligencia fluida (comprensión, semejanzas y toda la subescala manipulativa)⁸.

(1) Los mejores datos de los que se dispone, principalmente pruebas militares de varones jóvenes en Bélgica, Holanda e Israel, muestran ganancias de aproximadamente 20 puntos por generación (30 años), mientras que Noruega, Suiza y Dinamarca muestran unas ganancias de unos 10 puntos. Datos no tan ricos como los anteriores (Gran Betaña, Australia y Canadá) arrojan incrementos de entre 12 y 16 puntos por generación.

¿QUÉ IMPLICACIONES PRÁCTICAS TIENE EL EFECTO FLYNN EN EL USO DE LOS TESTS DE INTELIGENCIA?

Estos resultados, publicados por primera vez por Flynn en el *Psychological Bulletin*⁶ plantean cuestiones que todavía siguen generando grandes debates tanto en lo referente a la práctica de la medición de la inteligencia como en lo referente a la teoría subyacente. Analizaremos en primer lugar las consecuencias que el efecto Flynn tiene sobre la metodología de las pruebas de inteligencia.

La primera cuestión práctica que se deriva de los hallazgos de Flynn es obvia: si la principal utilidad de los tests es clasificar a un sujeto determinado con respecto a su población de referencia, y esta población está variando, ¿cómo van a cumplir entonces los tests su finalidad? Tomemos como ejemplo un estudio del propio Flynn: el error de medida para la mayoría de los tests es de aproximadamente 5 puntos, lo que significa que si un sujeto obtiene un resultado de 100 en una única aplicación de una prueba tenemos 19 posibilidades de 20 de que la puntuación real del sujeto esté en algún sitio entre 95 y 105. Pero si además ese sujeto no se corresponde con la generación de la muestra empleada para la baremación del test (lo que es bastante probable), aunque tuviéramos la medida real de su CI en puntuación directa, ésta podría traducirse dentro de un rango de entre 90 a 110 o incluso de entre 80 a 120, dependiendo del tipo de test. En resumen, la verdadera puntuación de un sujeto que ha obtenido un CI de 100 puede hallarse en cualquier lugar entre 70 y 130¹. Realmente esto es alarmante, aunque, como apunta Flynn², este efecto suele reducirse en la práctica profesional al introducirse el juicio clínico del evaluador. El mejor ejemplo para explicar cómo la variabilidad del CI influye en la precisión de los tests lo proporcionó el propio Jensen, quien propone comparar este fenómeno con un supuesto intento de medir la altura basándonos en la medida de las sombras proyectadas. Al variar el tamaño de las sombras en función de la posición del sol es prácticamente imposible conocer la altura real de un sujeto. Lo mismo sucede actualmente con los tests: al variar las puntuaciones de las poblaciones resulta imposible conocer la posición real de un sujeto determinado¹⁰.

No podemos olvidar, pues, el efecto Flynn a la hora de analizar los distintos estudios epidemiológicos o aquellos que comparan diferentes grupos y poblaciones en lo referente a las medidas de inteligencia. Valgan como ejemplo de las implicaciones del efecto Flynn sobre cualquier análisis poblacional los siguientes estudios documentados por Flynn: cita Flynn el realizado por Vernon en 1982, en el que se comparaba a un grupo de norteamericanos de raza china con un grupo de norteamericanos de raza blanca, obteniendo unos resultados que situaban a los norteamericanos de raza china dentro de una élite intelectual en comparación con sus conciudadanos de raza blanca. Pero lo que en realidad se estaba comparando era a un grupo de norteamericanos de raza china con

una generación anterior de norteamericanos de raza blanca; lo que daba cuenta de las diferencias intelectuales no era la raza, sino la diferencia generacional^{3,11}. En otro caso documentado por Flynn⁶, un grupo de puntuaciones de CI fue comparado con otro grupo, pero de una generación distinta, arrojando resultados incorrectos que otorgaban significativos incrementos de CI a niños adoptados por familias de más alto nivel socioeconómico al compararlos con los de sus madres biológicas. Siguiendo con esta lógica también se pueden poner en duda creencias acerca de las propiedades de la inteligencia: es una hipótesis muy sostenida el hecho de que la inteligencia fluida disminuye con la edad. Este fenómeno nace de comparar las puntuaciones de inteligencia de personas adultas con las puntuaciones de los jóvenes de hoy día. Sin embargo, si comparáramos las puntuaciones de esos adultos con las de los jóvenes de su generación (los de medio siglo antes) las pérdidas de CI serían prácticamente inapreciables¹².

Otro campo estrechamente relacionado con el uso de los tests de inteligencia es el de la evaluación de las personas con discapacidad intelectual⁽²⁾. Desde hace ya más de medio siglo la American Association on Mental Retardation (AAMR) ha sido la principal autoridad a la hora de establecer los criterios diagnósticos de la discapacidad intelectual. En todas sus ediciones de definición de retraso mental, 10 hasta la fecha¹³, uno de los criterios necesarios es un funcionamiento intelectual significativamente menor a la media, que en 1983 se consolidó como un CI de 70 o menor^{14 (3)}. ¿Cuál es la justificación de esta línea de corte y cómo afecta a la clasificación de las personas con discapacidad intelectual? Entre 1947 y 1948 se llevó a cabo la estandarización de la Weschler Intelligence Scale for Children (WISC) con una muestra representativa de los niños norteamericanos de la época. La media de ejecución se definió como un CI de 100; la ejecución en el percentil 16 (una desviación típica por debajo de la media) se definió como un CI de 85, y una ejecución en el percentil 2,27 (dos desviaciones típicas por debajo de la media) se definió como un CI de 70. En otras palabras, un CI de 70 o inferior representa el 2,27% de la población que se encuentra en el extremo inferior de la curva de inteligencia (esto no es del todo cierto, ya

que la mayoría de los tests de inteligencia sólo cuentan en su baremación con una muestra de población «biológicamente normal»; es decir, no incluyen a aquellos sujetos que sufren daños cerebrales o que presentan síndromes de etiología genética o cromosómica. De incluirlos, el porcentaje debería ascender a un 3%)¹⁵. Al considerarse que la problemática para el desarrollo de conductas adaptativas es característica de entre un 2 y 3% de la población gracias a las matemáticas de la curva normal, dos desviaciones típicas por debajo de la media seleccionan al 2,27% de una población «biológicamente normal»⁽⁴⁾. Es por ello que ya desde 1945 ha habido un consenso internacional para adoptar la puntuación de un CI de 70 o menos como el criterio de corte para establecer un diagnóstico de discapacidad intelectual.

Aquí topamos con el primer problema: el día después de la baremación del test las ganancias de CI van reduciendo ese famoso 2,27%. Además, en ocasiones, estas ganancias son mayores entre los sujetos con puntuaciones de CI menores de 70⁽⁵⁾.

Así pues, entre que se baremó la primera WISC entre 1947 y 1948 y la baremación de 1972 de la WISC-R, los niños con un CI de 70 habían incrementado sus puntuaciones en 8,25 puntos de CI¹⁶. Cada año que transcurría más y más niños sobrepasaban la puntuación de 70, llegando a que en 1972 un CI de 70 sólo recogía al 0,54% del final de la curva. Así, cuando en 1974 se publicó la WISC-R con la nueva estandarización, de la noche a la mañana el porcentaje de aquellos clasificables como sujetos con discapacidad intelectual remontó dramáticamente hasta volver a recoger al 2,27% del final de la curva bajo un CI de 70 (para inmediatamente volver de nuevo a recoger cada vez menos sujetos en la horquilla de 70 o menos). En resumen, las pruebas de Weschler nos dan una puntuación que supuestamente aísla a un 2,27% de la población «biológicamente normal», aunque esto es solamente momentáneo, pues la curva normal comienza a desplazarse de nuevo hacia la derecha, dejando cada vez menos sujetos por debajo de una puntuación de 70. Este fenómeno obviamente tiene serias repercusiones a la hora de clasificar a las personas con retraso mental: dependiendo de la baremación con la que comparemos a un sujeto determinado puede variar su tipo de retraso o incluso camuflar su existencia. Así pues, sólo podemos conocer el CI real de un sujeto si forma parte de la muestra con la que se ha baremado la prueba. De igual modo podemos utilizar el efecto Flynn con fines distintos a los del conocimiento de su funcionamiento intelectual: si queremos que un sujeto obtenga un diagnóstico de dis-

(2) Utilizaremos indistintamente a lo largo del artículo el término «discapacidad intelectual» y «retraso mental».

(3) Son tres los criterios establecidos por la AAMR para el diagnóstico del retraso mental: limitaciones significativas en el funcionamiento intelectual y la conducta adaptativa con un comienzo anterior a los 18 años. Cabe decir que las definiciones anteriores a la 9.ª edición de 1992 no consideraban las limitaciones de la conducta adaptativa como un criterio diagnóstico, tomando únicamente la medida del CI y la edad de comienzo inferior a los 18 años. Fue en 1992, con el revolucionario planteamiento que supuso esta nueva definición, cuando se incluyeron las limitaciones en la conducta adaptativa como criterio necesario, precisamente con la intención de eliminar el reduccionismo y la excesiva confianza en el uso de los test de inteligencia.

(4) Flynn ha intentado rastrear los datos aportados por Wechsler para la justificación de este 3% de la población, concluyendo que estos datos no parecen existir.

(5) Flynn ha analizado cuidadosamente las diferencias en las ganancias con el paso del tiempo entre la muestra por debajo de un CI de 70 y el resto. Dichas diferencias son significativas, lo que plantea dudas acerca de la medición del retraso mental.

capacidad intelectual para así conseguir una mayor calificación de minusvalía o alguna prestación económica aplicaremos una prueba de baremación reciente, mientras que si lo que buscamos es ocultar posibles disfunciones intelectuales aplicaremos una prueba baremada con generaciones pasadas.

Pero todo esto también tiene serias repercusiones en lo referente a los distintos estudios epidemiológicos del retraso mental. Este tipo de estudios son recientes en el campo de la discapacidad intelectual, aunque cada vez más numerosos¹⁷. Intentan dirimir principalmente la prevalencia e incidencia del retraso mental y de sus distintos grados. En la mayoría de ellos se toman las medidas de las pruebas de Weschler y se analizan los datos para ver las diferencias en las tasas de retraso mental a lo largo del tiempo. Con lo repasado hasta ahora es lógico concluir que si no se tiene presente el efecto Flynn los resultados de estos estudios poca luz podrán arrojar en lo referente a los cambios en la prevalencia e incidencia de la discapacidad intelectual.

¿Qué hacer entonces para solventar todos estos problemas prácticos que presenta la utilización de los tests de inteligencia en lo que se refiere al efecto Flynn? Antes de nada decir que pese a estas cuestiones prácticas el trabajo detrás de la elaboración de cada prueba de inteligencia es admirable en cuanto a su seriedad y magnitud y que los tests han supuesto un importante empuje para la comprensión de la inteligencia y más concretamente para la población con retraso mental. Sin embargo, sigue habiendo ciertos temas que deben estudiarse para continuar mejorando la fiabilidad y validez de este tipo de pruebas. Con respecto al control del efecto Flynn, una posible solución sería la de calcular el patrón de incremento para cada prueba e ir aplicándolo para ajustar las medidas en función de la fecha en la que han sido tomadas. Desgraciadamente, la ratio de incremento del CI es demasiado variable para poder establecer un patrón constante. Por ejemplo, cuando se baremó la WISC-R en 1972 usando las ratios de los años anteriores se hubieran previsto unas ganancias para 1989 de unos 5 puntos ($0,3 \times 17$ años). La baremación de la WISC-III en 1989 sí mostró estas ganancias de 5 puntos, pero, sin embargo, la baremación de la *Weschler Adults Intelligence Scale III* (WAIS-III), también con 17 años de diferencia mostró ganancias de sólo 3 puntos¹. Quién sabe si algún día cuando se cuente con más datos de estas variaciones del CI se pueda establecer un patrón más detallado para cada tipo de prueba en cada población.

Otra solución propuesta por Flynn es la de abandonar los tests de inteligencia por escalas de observación conductual que midan la conducta adaptativa⁹. El problema de esta opción es que quizás estas medidas también vengán acompañadas de algún fenómeno similar al de Flynn, en cuyo caso no habríamos avanzado nada; más bien habríamos perdido todo el cuerpo de conocimiento que han generado los tests de inteligencia. Un segundo problema que plantea esta solución propuesta por Flynn es la pérdida de la perspectiva etiológica. Bajas puntuaciones en una escala de habilidades de adaptación también

pueden deberse a un trastorno esquizofrénico, una depresión o a un entorno familiar desestructurado. Con todo esto quizá la solución no sea cambiar los tests de inteligencia por medidas de observación conductual, sino, como propone la AAMR¹³, añadir las como un instrumento más, necesario para cualquier clasificación intelectual.

Asimismo cabe citar en este apartado las investigaciones de Jensen respecto a posibles variables fisiológicas que muestren correlaciones con el desempeño intelectual. Los experimentos de Jensen sobre la respuesta eléctrica del córtex cerebral, sobre los tiempos de reacción o sobre los tiempos de absorción de glucosa por el cerebro¹⁸ pueden parecer un futuro prometedor. Y quizá lo sea en lo que se refiere a la búsqueda de una medida libre de sesgos culturales, pero, como sucede con la propuesta de Flynn de medidas de observación conductual, quién sabe si este tipo de medidas que propone Jensen también se encuentran afectadas por fenómenos similares al efecto Flynn.

Por último, la única solución restante parece ser la de baremar continuamente cada prueba de inteligencia para cada población. Flynn propone una reestandarización cada 7 años para conferir un mínimo de respetabilidad al CI, aunque duda acerca de si una baremación tan periódica sea tan gravosa que no sea práctica⁹. Si bien es cierto que esta opción es costosa y que además puede tener efectos colaterales que revisaremos más adelante, parece un buen método para minimizar las consecuencias que los progresivos incrementos del CI tienen sobre la precisión de las pruebas de inteligencia.

¿QUÉ IMPLICACIONES TEÓRICAS TIENE EL EFECTO FLYNN EN EL USO DE LOS TESTS DE INTELIGENCIA?

Después de este breve recorrido por las consecuencias que el efecto Flynn puede tener sobre la utilización de las pruebas de inteligencia, pasamos ahora a centrarnos en el arduo debate teórico que los resultados de Flynn despiertan acerca de las medidas de inteligencia. En cuanto a las dudas teóricas que plantea el efecto Flynn, la primera, obviamente, es la que se refiere a si estos incrementos del CI son ganancias reales de lo que entendemos como inteligencia o si están dando cuenta de otro fenómeno subyacente. Retomemos un nuevo ejemplo práctico de las consecuencias del efecto Flynn para ilustrar esta cuestión: si comparamos los datos obtenidos por John Raven de dos cohortes de entre 25 y 65 años, una medida en 1942 y otra en 1992 y comparamos sus puntuaciones directas concluimos que el 90% de los nacidos en 1877 se sitúan por debajo del percentil 5 de los nacidos en 1967, es decir, por debajo de un CI de 75 calculado en 1967¹⁹. ¿Realmente el 90% de la población nacida en 1877 tenía retraso mental si la observamos con los baremos actuales? Otro ejemplo esclarecedor es el ya comentado respecto a las fluctuaciones en el porcentaje de personas con discapacidad intelectual, es decir, de personas que se sitúan por debajo de un CI de 70. Si el incremen-

to de las puntuaciones del CI fuera realmente un incremento de la inteligencia, entonces el número de personas con retraso mental hubiera realmente disminuido. Pero de haber sido real esta disminución, la publicación de la WISC-R de 1974, por ejemplo, tendría que haber puesto en pie a todos los profesionales del ámbito del retraso mental por haber vuelto a abarcar a un 2,27 % de la población dentro del criterio de discapacidad intelectual.

Con todo esto, ¿cada generación es más inteligente que la anterior, están actuando unos factores subyacentes que dan cuenta de este fenómeno y que son independientes de la capacidad intelectual o se trata de una mezcla de ambos?

Hasta la fecha se han barajado numerosísimas hipótesis que exponemos a continuación. Normalmente se presentan todas juntas, pero creemos que hay una distinción importante que debe hacerse entre aquellas que dan a entender que se están dando incrementos reales en lo que entendemos como inteligencia y aquellas que apoyan que los incrementos de Flynn no reflejan ganancias reales de inteligencia.

Entre las hipótesis que apuntan a que los incrementos del CI reflejan incrementos en la inteligencia de la población cabe destacar las siguientes:

Mejoras en la nutrición. Es claro que en el último siglo se han experimentado mejoras en la nutrición, pero no es tan clara la relación que esto pueda tener con incrementos en la inteligencia. Richard Lynn es el principal defensor de esta postura, afirmando que ella sola da cuenta del efecto Flynn: cerebros mayores producen mayores niveles de inteligencia. Los incrementos de CI, pues, son incrementos reales de inteligencia. Existen datos suficientes que avalan la correlación existente entre malnutrición y bajas puntuaciones de CI, pero no hay evidencia de lo contrario, a saber: que una mejor nutrición se correlacione con una mejor inteligencia. Con respecto a este punto, recordemos que se encuentran ganancias en todos los estratos de la curva de inteligencia, no sólo en aquellos que están por debajo de la media. Tampoco se han hallado correlaciones entre el incremento de la estatura y el incremento del CI. Y los patrones de incremento de inteligencia tampoco se han visto afectados por las guerras o hambrunas³.

Nivel socioeconómico y urbanismo. Se trata de hipótesis lógicas, pues es obvio que el nivel socioeconómico ha experimentado un aumento y que el urbanismo ha mejorado notablemente incrementando la comunicación y disminuyendo el aislamiento. Sin embargo, los datos existentes plantean dudas, ya que las correlaciones entre incremento de CI e incremento de nivel socioeconómico no son tan altas como para dar cuenta del total del efecto Flynn³. Lo mismo sucede con los estudios que reflejan los efectos de la urbanización de las poblaciones: los cálculos del déficit cognitivo de las zonas rurales no pueden dar cuenta por sí solos de los incrementos encontrados por Flynn³.

Hipótesis de Storfer. Es una hipótesis similar a la de Lynn, pero pone el peso en la erradicación de las enfermedades infantiles y en las mejoras de la calidad cogniti-

va de los ambientes preescolares y del hogar²⁰. También podría entenderse como una consecuencia de las mejoras urbanísticas y socioeconómicas, por lo que todo lo dicho sobre las dos hipótesis anteriores podría aplicarse a los planteamientos de Storfer.

Educación. El aumento de años de escolarización y la mejora de los métodos y contenidos pedagógicos resulta una causa muy probable para dar cuenta de los incrementos del CI. En cuanto al aumento de los años de escolarización, en todos y cada uno de los 20 países empleados en los estudios de Flynn se han incrementado los años de escolaridad. Así, por ejemplo, las ganancias de CI de Dinamarca obtienen altas correlaciones con el incremento de los años de escolarización²¹. Sin embargo, en Holanda un estudio de emparejamiento intergeneracional para mantener constante el nivel educativo obtuvo que éste sólo daba cuenta del 6,5 % de las ganancias observadas en dicho país⁵.

En lo referente a la mejora de los métodos y contenidos pedagógicos, los resultados observados también plantean problemas. Un estudio llevado a cabo en Holanda en 1982 por Rist³ demostró que los estudiantes educados con los nuevos métodos de pedagogía matemática al llegar a la edad militar y ser evaluados en los subtests de aritmética de la WAIS no sólo no presentaban ganancias del CI en este aspecto, sino que presentaban pérdidas.

Pero dentro de este apartado habría que hacer una nueva distinción: ¿ha mejorado el contenido del currículum académico o ha mejorado la enseñanza de habilidades abstractas de resolución de problemas? Como expusimos al comienzo, la mayor parte de las ganancias en el CI se debe a las pruebas de medidas fluidas, encontrándose pocos incrementos en las pruebas cristalizadas. Esto parece dejar claro que de haber una mejora pedagógica susceptible de reflejarse en ganancias de CI ésta tiene que haberse dado en lo que se refiere a resolución de problemas. Pero, ¿cómo identificar esas prácticas pedagógicas? ¿Cómo medirlas? ¿Están realmente relacionadas con los tests de inteligencia fluida?

Como puede deducirse, la hipótesis de la educación entraña complicados planteamientos. Para acabar citaremos el estudio de Cahan y Cohen de 1987 en el que comparan los efectos de un año escolar (controlando la edad) con los de un año cronológico (controlando los años de escolarización) en 10.000 niños de colegios de Jerusalén. Como era de esperar daba más cuenta de la varianza la escolarización que la edad, pero, sorprendentemente, ésta afectaba más a las aptitudes de inteligencia fluida que de inteligencia cristalizada²².

Nuevas prácticas de crianza. Con la modernización de las sociedades se han experimentado cambios en las prácticas de crianza. Los padres están interesados en el desarrollo intelectual de sus hijos y les estimulan en este sentido. Proliferan los libros sobre consejos para las buenas prácticas de crianza de los hijos, las investigaciones sobre el tema (Spock, Bowlby, etc.), así como los programas televisivos pedagógicos. Pero el efecto que todo esto pueda tener sobre el desarrollo del CI de los infantes

no está claro. Es sabido que los programas preescolares como Head Start no producen cambios duraderos del CI. Sin embargo, no podemos descartar que intervenciones más tempranas y continuadas sí los tengan²². Por ejemplo, al comparar a los 57 niños con una edad media de 4 años que se beneficiaron del programa Abecedarian Project en Carolina del Norte con niños de un grupo control, se encontraron diferencias significativas que no disminuyeron con el tiempo. Sin embargo, esta intervención sólo resultó en ganancias de 5 puntos del CI, lo que dista mucho de acercarse a las ganancias halladas por Flynn.

Ambiente más técnico y visual. Esta hipótesis, propuesta por Neisser²², engloba lo que ya ha sido apuntado por distintos autores: en todos los países estudiados, todos ellos de civilizaciones que podríamos denominar como más «occidentalizadas», se han producido enormes cambios⁽⁶⁾. Se han producido cambios en las aspiraciones, los modelos, la información, el ocio, etc. Destaca Neisser el auge de los medios visuales: pósters, *graffiti*, cine, televisión, videojuegos, ordenadores, etc., y argumenta que los niños expuestos a estos medios desarrollan habilidades visuales específicas. Existen pocos datos para el análisis de esta hipótesis, pero la recogemos por su alto poder intuitivo. En este sentido también apuntan las hipótesis de Vincent²³. Dicho autor cree que la complejidad del mundo moderno es responsable y da cuenta en sí misma de las ganancias del CI. Los estímulos diarios a los que estamos sometidos desde la revolución industrial son suficientes para justificar esos masivos incrementos de inteligencia.

Reducción de los niveles de exposición al plomo. Nuestro grupo ha estimado recientemente las pérdidas en el CI en las poblaciones vinculadas con la exposición al plomo en el ambiente²⁴. Se puede llegar a producir una reducción de 2,6 puntos de CI por cada incremento de plomo en sangre de entre 10 a 20 µg/dl, asumiendo una pérdida de 3,5 puntos de CI en niveles superiores a 20 µg/dl²⁵. Las medidas de plomo en sangre obtenidas para este estudio citado muestran niveles de plomo en la actualidad que oscilan entre los 1,7 µg/dl y los 15,4 µg/dl, la primera correspondiente a la media de Estados Unidos y Canadá y la segunda a Egipto, Marruecos y Pakistán. Un repaso minucioso de estas medidas de distintos países ofrece un patrón de reducción de plomo similar al patrón de «occidentalización». Por todo ello podemos decir que la disminución de los niveles ambientales de plomo como consecuencia de la implementación de medidas gubernamentales que favorecen el uso de gasolina sin plomo que ha tenido lugar en los países desarrollados y en muchos de los países en vías de desarrollo puede dar cuenta de parte de los incrementos del CI, aunque no del total de las ganancias.

Acabamos de hacer un repaso por las distintas hipótesis que parecen indicar que los incrementos del CI hallados por Flynn son incrementos reales de la inteligencia (mejoras en la nutrición, mejoras del nivel socioeconómico

y del urbanismo, la hipótesis de Storfer, las nuevas prácticas de crianza, la reducción de los niveles de exposición al plomo). Otras, como el ambiente más técnico y visual y un enfoque distinto de los efectos de la educación, pueden entenderse como hipótesis mixtas, en el sentido de que también pueden reflejar otras variables que influyan en el resultado de los tests, pero que no afecten directamente a la inteligencia. A continuación exponemos las hipótesis más extremas en este aspecto, es decir, aquellas que justificarían los incrementos del CI, pero que no derivarían en el hecho de que estos incrementos fueran reales.

Hipótesis de Brand. Brand^{26,27} argumenta en su hipótesis cómo el aumento de los niveles de permisividad en las sociedades afecta a un mejor desempeño en las pruebas de inteligencia. Los incrementos de CI se correlacionan con aumentos en la promiscuidad, divorcio, falta de creencias religiosas, consumo de tabaco, accidentes, crimen, pues ambos fenómenos son consecuencia de este aumento de permisividad. La falta de meticulosidad y el desprecio a las reglas y falta de miedo a las consecuencias hacen que se responda a los tests sin perder tiempo ante las preguntas complicadas y que aumente el número de respuestas al azar, lo que beneficia al desempeño aunque no a la inteligencia. Raven²⁸ aporta datos que desdicen en parte esta hipótesis al argüir que el efecto de las respuestas al azar no incrementa la ejecución del test. De igual modo, al comparar dos generaciones en el test de Binet encontraron que era precisamente la segunda generación la que más ítems dejaba sin contestar³.

Sofisticación de los tests. Esta hipótesis se refiere al incremento de la utilización de tests en el siglo XX. Hemos hecho una traducción literal del nombre con el que aparece en los distintos artículos, aunque quizá fuera más ilustrativo nombrarla como algo parecido al aumento de la presencia de los tests en la sociedad. Cada vez estamos más familiarizados con el uso de los tests de inteligencia y cada vez tenemos más práctica en responder a sus preguntas. Todo esto, indudablemente, deberá afectar a su ejecución; la duda está en el grado de afectación. Flynn no puede contradecir este efecto, aunque intenta minimizarlo afirmando que los incrementos del CI no se han visto afectados por las fluctuaciones en la popularidad de los tests³. También afirma en este sentido que los sujetos a los que se les administra repetidamente un test sólo ofrecen ganancias de 5 ó 6 puntos de CI.

EL PRINCIPIO DE INCERTIDUMBRE DE FLYNN

Es precisamente en esta última hipótesis donde discrepamos de lo dicho hasta ahora. En el momento en el que los tests fueron introducidos, la sociedad se familiarizó con ellos, conoció sus reglas y la forma de pensar subyacente a las pruebas de inteligencia caló en padres, pedagogos y estudiantes (no hay más que analizar los programas de desarrollo de inteligencia, que en su mayoría son entrenamientos en las habilidades necesarias para responder correctamente a los tests). Es por ello

⁽⁶⁾ Las muestras obtenidas en China provienen exclusivamente de las zonas urbanas.

que la popularidad de los tests no debe en teoría afectar, porque no es el número de aplicaciones lo que mejora el desempeño, sino la forma de razonamiento que proponen; una vez que esta forma de razonamiento aparece en una cultura, su efecto multiplicador no puede detenerse, independientemente del número de veces que se apliquen las pruebas. Del mismo modo, el efecto de la presencia de los tests nada tiene que ver con una doble aplicación de una prueba a un mismo sujeto. Ese sujeto ya ha nacido en un ambiente donde los tests están presentes, y de eso dependerá su ejecución. El tipo de razonamiento de un test ya lo conoce, por lo que la información que obtenga al realizar una prueba será, en su gran mayoría, redundante. Tomaremos un ejemplo del mundo del deporte para ilustrar este punto: un corredor de 100 metros lisos nacido en 1992 obtendrá unos tiempos mucho menores que un corredor de su misma edad pero nacido 50 años antes. Más aún: ambos corredores tienen la misma forma física ¿A qué se debe entonces esta diferencia en la ejecución? Lógicamente a los conocimientos acumulados sobre las técnicas de atletismo. Desde que se instauraron los 100 metros lisos como prueba, y desde que ésta comenzó a tener popularidad, cada avance en el conocimiento de las técnicas de entrenamiento, dietas alimentarias, concentraciones, etc., se ha ido acumulando y cada nuevo corredor no parte de cero, sino que se beneficia de estos conocimientos de una cultura que ya lleva un siglo conociendo las carreras de 100 metros lisos. Lo mismo sucede con cualquier deporte: sólo nos basta con observar cómo se jugaba al tenis hace 50 años y cómo se juega ahora, ver el *swing* de golf de un jugador de hace 50 años o el de los jugadores actuales, un partido de fútbol de entonces o un partido actual, partidos de baloncesto, etc. Pero lo que sí parece claro es que estas mejoras se deben al conocimiento acumulado sobre cada deporte y no a una mejora en las capacidades innatas de los deportistas. Quizás algo se deba a las mejoras en la nutrición, por ejemplo; pero parece claro que éstas dan menos cuenta de las ganancias que los conocimientos deportivos. Cojamos ahora a un jugador de tenis nacido en 1877 que se sitúa entre los 10 mejores y pongámoslo a jugar en una liga de jugadores nacidos en 1967. ¿Qué sucedería? Aunque los igualáramos en su forma física probablemente nuestro jugador de 1877 se situaría entre los 30 peores, pero seguramente debido a los conocimientos acumulados sobre las técnicas del tenis que posee la **generación de 1967**⁽⁷⁾.

Por eso, visto desde esta perspectiva, no es válido el argumento de que al pasar un test dos veces a un mismo sujeto su ejecución sólo mejora en 5 ó 6 puntos. Un jugador de tenis nacido en 1967 que juegue dos partidos consecutivos no mejora de uno a otro.

⁽⁷⁾ Todo esto alude a otro debate, abordado por Flynn y por los teóricos de la inteligencia que aunque no trataremos en este artículo, sí querríamos señalar, a saber: ¿qué miden realmente los tests de inteligencia fluida, el potencial puro de un sujeto o el desempeño de ese potencial?

Un ejemplo directo del campo de los tests y que nos servirá más adelante para dar cuenta de las diferencias halladas dentro de las ganancias de CI es el de la propia prueba de las Matrices progresivas de Raven. Es un fenómeno conocido por todos que si abrimos el cuadernillo de Raven por sus últimas páginas y nos topamos, por ejemplo, con el ítem D-7, nos resultará complicado deducir que la respuesta correcta es la 5. Sin embargo, si hacemos el test en su orden tardaremos poco tiempo en contestar correctamente al ítem D-7 porque hemos ido acumulando un conocimiento sobre qué se nos está pidiendo en la prueba. Es más, Carpenter, Just y Shell²⁹ han identificado las cinco reglas necesarias para completar las matrices y que se van adquiriendo a lo largo de la prueba: *a)* existe un patrón numérico entre las matrices adyacentes; *b)* el mismo valor se mantiene en las filas pero cambia en las columnas; *c)* una fila de una columna sumada o restada a otra produce la tercera; *d)* tres formas distribuidas a través de una fila siempre deben estar presentes, y *e)* dos valores están distribuidos en una fila y un tercer valor es nulo. Éstas son, pues, las cinco reglas que vamos descubriendo a través del Raven. En el momento en el que desciframos una acumulamos ese conocimiento, mejorando la ejecución.

Desde esta nueva perspectiva analicemos ahora los distintos datos que ha ido presentando Flynn. El dato más importante aportado hasta la fecha parece ser el de mayores tasas de incremento en los tests de inteligencia fluida que en los de inteligencia cristalizada. Los conocimientos que se evalúan a través de estos últimos han estado siempre presentes, incluso antes de la aparición de los tests de inteligencia. Recordemos que las subescalas de información, dígitos y aritmética apenas han sufrido incrementos (incluso han presentado pérdidas en algunos casos: información -0,3; dígitos +0,1 y aritmética +0,3)⁹. La aparición de los tests no ha podido influir en los conocimientos que se tienen sobre la cultura, la memoria numérica o las matemáticas. Lo mismo sucede con el subtest de vocabulario, con un incremento medio de +0,4. Sin embargo, el resto de los subtests (comprensión, figuras incompletas, rompecabezas, clave de números, historietas, cubos y semejanzas) que han presentado ganancias de 2 a 6 veces mayores sí requieren habilidades no presentes habitualmente. Pero hay un dato todavía más interesante: en las cuatro pruebas que apenas muestran incrementos, aunque sus ítems estén presentados en dificultad creciente, el orden de respuesta no debía afectar al resultado. Sin embargo, en el resto de subtests sí **afecta**⁽⁸⁾. Preguntar quién era Gandhi como ítem 20 o como ítem 1 no afecta al resultado, pues se conoce la respuesta o no se conoce. Repetir la secuencia 4-2-7-3-1 antes o después de una secuencia de menos guarismos tampoco afecta el resultado, pues se tiene la memoria suficiente o no se tiene.

⁽⁸⁾ Quizás el único subtest cuestionable en este aspecto sea el de comprensión, aunque cabe decir que después de los cuatro subtests mencionados como aquellos que menos ganancias presentan el siguiente es precisamente el de comprensión.

Proponer una división antes o después que una resta no afecta a la ejecución, pues depende del manejo previo que se tenga de las distintas operaciones matemáticas. Por último, responder a qué significa la palabra «ominoso» en la pregunta 33 o antes tampoco afecta, pues se conoce su significado o no se conoce. Sin embargo, en el resto de subtests sí influye la dificultad creciente, pues requieren habilidades no presentes habitualmente, de ahí que el entrenamiento que supone el incremento de dificultad influya en la **ejecución**⁽⁹⁾. Creemos que este pequeño análisis alumbrará que las ganancias en los tests de inteligencia se dan en aquellas habilidades de razonamiento no presentes habitualmente y que han sido introducidas con la aparición de los tests de inteligencia.

Otro dato de interés es el aportado por Flynn respecto a que existen indicios de que la escalada de ganancias de CI llegará a su fin a finales del siglo XX o principios del XXI, incluso en las pruebas de inteligencia fluida⁸. Si mantenemos nuestra hipótesis de este nuevo conocimiento que han traído consigo los tests, no es de extrañar que, efectivamente, éste llegue a un tope algún día por una saturación de su presencia, como ya ha sucedido con los tests de inteligencia cristalizada, aproximadamente en 1948⁸.

Pero, ¿qué implicaciones tiene esta nueva hipótesis en el uso de los tests? Hemos visto cómo la introducción de los tests ha supuesto, precisamente, una escalada en las habilidades necesarias para la cumplimentación de las pruebas de inteligencia, principalmente las de inteligencia fluida. Es decir, que la introducción de los tests ha provocado una alteración de aquello que pretenden medir. Recordemos el principio de incertidumbre de Heisenberg: hacia el año 1926 el físico Werner Heisenberg formuló el principio de incertidumbre o indeterminación fundamentándose en un experimento en el cual se intentaban medir dos variables: el momento y la posición del electrón. Como una partícula no se puede observar directamente había que realizar un análisis estadístico y así obtener una «probabilidad» resultante de la medición. Pero lo más sorprendente de este experimento es que cuando se intentaba medir la posición ya no podía determinarse el momento, y lo mismo sucedía si se intentaba medir el momento, pues ya no podía determinarse la posición. Así pues, la física cuántica nos dio a conocer que al realizar una medición se cambia el comportamiento de las partículas sometidas a esa medición. Algo parecido sucede con nuestra hipótesis y el efecto Flynn: al utilizar un test como instrumento de medida estamos alterando aquello que queremos medir.

Hemos hecho un breve repaso de los datos de Flynn, de las consecuencias prácticas que tienen para el uso de los tests y las posibles soluciones, de los debates teóricos que despierta y las distintas hipótesis explicativas,

para acabar centrándonos en la hipótesis que consideramos más factible. No queremos acabar sin decir que no descartamos que entre las ganancias halladas por Flynn también pueda estar presente una conjunción de todas las causas antes repasadas. Indudablemente el debate está abierto y cada vez son más los intentos de hallar exhaustivos modelos explicativos de este fenómeno que se conoce como efecto Flynn.

AGRADECIMIENTOS

Trabajo llevado a cabo dentro del proyecto FIS G03/061.

BIBLIOGRAFÍA

1. Flynn JR. The schools: IQ tests, labels, and the word intelligence. En: Carlson JS, Kingma J, Tomic W, editores. *Advances in cognition and educational practice*. Vol. 5. Conceptual issues in research on intelligence. Londres: JAI Press, 1998; p. 13-42.
2. Herrnstein RJ, Murray C. *ART 5. The bell curve: intelligence and class in american life*. New York: Free Press, 1994.
3. Flynn JR. IQ gains over time: toward finding the causes. En: Neisser U, editor. *The rising curve: long-term gains in IQ and related measures*. Washington: American Psychological Association, 1998; p. 25-66.
4. Horn JL. Cognitive diversity: a framework for learning. En: Ackerman PL, Sternberg RJ, Glasser R, editores. *Learning and individual differences: Advances in theory and research*. New York: Freeman, 1989; p. 61-114.
5. Flynn JR. Massive IQ gains in 14 nations: what IQ tests really measure. *Psychol Bull* 1987;101:171-91.
6. Flynn JR. The mean of IQ of americans: massive gains 1932 to 1978. *Psychol Bull* 1984;95:29-51.
7. Weschler D. *WISC-III: Weschler Intelligence Scale for children*, 3rd ed. Manual. San Antonio: The Psychological Corporation, 1992.
8. Flynn JR. IQ gains, WISC subtests and fluid g: g theory and the relevance of Spearman's hypothesis to race. En: Bock GR, Goode JA, Webb K, editores. *The nature of intelligence: symposium on the nature of intelligence*. Novartis Foundation. Londres: John Wiley and Sons, 2000; p. 202-27.
9. Flynn JR. The hidden history of IQ and special education: can the problems be solved? *Psychology, Public Policy, and Law* 2000;6:191-8.
10. Jensen AR. Commentary: vehicles of g. *Psychol Sci* 1992; 3:275-8.
11. Flynn JR. Asian americans: achievement beyond IQ. En: Hillsdale NJ, editor. Erlbaum, 1991.
12. Raven JC, Court JH, Raven J. *Manual for Raven's progressive matrices and vocabulary scales*. Oxford: Oxford Psychologist Press, 1992.
13. Luckasson R, Borthwick-Duffy S, Buntix WHE, Coulter DL, Craig EM, Reeve A, et al. *Mental retardation. Definition, classification and systems of supports*, 10th ed. Washington: American Association on Mental Retardation, 2002.
14. Grossman HJ, editor. *Classification in mental retardation*. Washington: American Association on Mental Deficiency, 1983.
15. Jensen AR. *Bias in mental testing*. Londres: Methuen, 1980.

⁽⁹⁾ Esto es cierto para todos los subtests menos para el de clave de números. Sin embargo, éste cuenta exclusivamente con una variable que también ha sido introducida por los tests de inteligencia: el tiempo límite de ejecución.

16. Flynn JR. Weschler Intelligence Tests: do we really have a criterion of mental retardation? *Am J Mental Defic* 1985; 90:236-44.
17. Yeargin-Allsopp M, Boyle C. Overview: the epidemiology of neurodevelopmental disorders. *Ment Retard Dev Disabil Res Rev* 2002;8(3):113-6.
18. Jensen AR. Rising IQ without increasing g? A review of the Milwaukee Project: preventing mental retardation in children at risk. *Develop Rev* 1989;9:234-58.
19. Flynn JR. Searching for Justice: the discovery of IQ gains over time. *Am Psychol* 1999;54:5-20.
20. Storfer MD. Intelligence and giftedness: the contributions of heredity and early environment. San Francisco: Jossey-Bass, 1990.
21. Teasdale TW, Owen DR. Continued secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence* 1989;13:255-62.
22. Neisser U. Rising Scores on Intelligence Tests. *Am Sci*, 1997.
23. Vincent KR. On the perfectibility of the human species: evidence using fixed reference groups. *Texas Counselling Assoc J* 1993;22:60-4.
24. Fewtrell LJ, Prüss-Ústün A, Landrigan P, Ayuso-Mateos JL. Estimating the global burden of disease of mild mental retardation and cardiovascular diseases from environmental lead exposure. *Environmental Research* 2004;94:120-33.
25. Fewtrell LJ, Fewtrell L, Prüss A, Landrigan P, Ayuso Mateos JL. Burden from environmental lead exposure. En: Murray C, López A, Ezzati M, editores. *Quantifying global health risks: the burden of disease attributable to selected risk factors*. Geneva: World Health Organization, 2003 (en prensa).
26. Brand CR. Bryter still and bryter? *Nature* 1987;328:110.
27. Brand CR. Keeping up with the times. *Nature* 1987;328:761.
28. Raven J. Methodological problems with the 1992 standardization of the SPM: a response. *Personality and Individual Differences* 1995;18:443-5.
29. Carpenter PA, Just MA, Shell P. What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychol Rev* 1990;97:404-31.