

Test de Raven: Actualización de Baremos en Adolescentes Argentinos y Análisis del Efecto Flynn

Raven's Progressive Matrices: Update Norms in Argentina and Analysis Flynn Effect

Lilia Rossi-Casé¹, Rosa Neer¹, Susana Lopetegui¹, Stella Doná¹, Bruno Biganzoli¹ y Ramiro Garzaniti¹

Resumen

Se presentan los resultados obtenidos en la construcción de normas actualizadas del Test de Matrices Progresivas de Raven, Escala General, para la ciudad de La Plata, Argentina; y su comparación con las obtenidas en los años 1964 y 2000. Con ese propósito se seleccionó una muestra de 1049 sujetos de ambos sexos, de 13 a 18 años de edad, alumnos de educación secundaria. El instrumento fue administrado colectivamente y sin límite de tiempo. Los resultados se analizaron en tres grupos de edad y se compararon con los baremos anteriores. Se observó que a mayor edad de los sujetos, su rendimiento es mejor. La variabilidad de los puntajes es similar en los grupos de edad. En relación con la tipificación del año 2000, en la que se constató el efecto Flynn respecto de las normas de 1964, aparece un amesetamiento y leve descenso de los puntajes directos necesarios para alcanzar el rendimiento promedio. Asimismo se observaron diferencias en el orden de dificultad entre las series de la prueba. El conjunto de estos resultados confirman la necesidad de actualización de los baremos y de las características del material de prueba, a los fines de una evaluación precisa.

Palabras clave: efecto Flynn, Test de Raven, baremos, adolescentes

Abstract

The current paper presents the results of the construction of valid and updated norms for Raven's Standard Progressive Matrices, in the city of La Plata, Argentina, and compares them with norms obtained in the years 1964 and 2000 in order to observe score trends. For this purpose a sample of 1049 subjects of both sexes, 13-18 years old, high school students were selected. The test was administered collectively, without time limit. Data was split into three age groups and results were compared with norms from 1964 and 2000. It was also noticed that as age increases, subjects performance tend to improve. The scores variability is similar in all age groups. A slight decrease of the direct scores needed to reach an average performance was observed, when compared to those from 2000, which were the norms where the Flynn effect was first noted after 1964 and show a plateau effect in direct scores. It was also noticed that as age increases, subjects performance tend to improve. The scores variability is similar in all age groups. Also differences in the order of difficulty between test series were observed. All these results confirm the need for updating of the scales and the characteristics of the test material for the purposes of an accurate assessment.

Keywords: Flynn effect, Raven's Progressive Matrices, norms, adolescents

¹Facultad de Psicología. Universidad Nacional de La Plata, La Plata, Argentina. Calle 51 entre 123 y 124, 3er. Piso, Secretaría de Investigaciones. Facultad de Psicología. Ensenada. Provincia de Buenos Aires. Argentina. (C.P. 1925). Teléfonos: +54 (0221) 482-4415 | 482-5931 | 482-8457 | 483-3933 | 483-4601 | 483-4871. Correo electrónico: rossicase@psico.unlp.edu.ar

Introducción

En este trabajo se presentan datos actualizados para la elaboración de normas de la Escala General del Test de Matices Progresivas de Raven. Surge principalmente del interés suscitado por una serie de fenómenos observados en las puntuaciones obtenidas por muestras que corresponden a diferentes generaciones, tanto a nivel internacional como en nuestro país. Complementariamente, de las características que pueden observarse en el propio instrumento respecto del orden de dificultad creciente de los problemas propuestos. Se trata en el primer caso del conocido Efecto Flynn y las controversias que se han suscitado acerca de su persistencia o eventual desaparición. En el segundo, de constataciones que emergen de nuestro trabajo con adolescentes escolarizados, en los que se pone en evidencia una graduación diferente del grado de dificultad de las series, establecido en los manuales de la prueba.

El Test de Matrices Progresivas de Raven

Sucintamente puede recordarse que es un test no verbal que examina la habilidad mental general de acuerdo a la teoría factorial de la inteligencia. Se presenta en varias formas: Escala General, Escala Paralela, Escala Especial Coloreada y Escalas Avanzadas. En la Escala General su administración puede ser individual o colectiva; es por ello autoadministrable, sin límite de tiempo en su versión original. La tarea del sujeto consiste en completar 60 matrices lacunarias distribuidas en 5 series de 12 ítems cada una, a partir de elegir la respuesta correcta de entre seis u ocho alternativas. La serie completa “integra una escala de matrices en orden de complejidad creciente construida de manera de revisar en la forma más completa posible los sistemas de pensamiento del desarrollo intelectual” (Raven, J. C., Raven, J., & Court, 2003, p. 9). Pone en juego procesos de *edución* de relaciones y correlaciones sobre un material en el que las variables a considerar no son obvias, es decir que se deben extraer nuevas comprensiones a partir de la información dada. Para Spearman, todo acto de conocimiento se rige por la *edución* (Raven, J. C., Raven, J. et al., 2003).

Fue publicado por primera vez en Gran Bretaña, en el año 1938 por su autor, John C. Raven, alumno del psicólogo inglés Charles Spearman, quien había enunciado en 1904 su “teoría ecléctica de los dos factores”. La *edución*, como operación cognoscitiva básica, se entiende a partir de las tres leyes en que este autor descompone el proceso de formación del conocimiento, conocidas como *leyes no genéticas*. La primera de ellas, que indica que ante dos o más ítems toda persona tiende a establecer relaciones entre ellas, se denomina ley de *edución de relaciones* (Raven, J. C., Raven, J. et al., 2003). La segunda, llamada de *edución de correlatos*, establece que ante un ítem y una relación, toda persona tiende a concebir el ítem correlativo. La tercera, según la cual toda persona tiende a conocerse de un modo inmediato a sí misma y a los ítems de su propia experiencia, se conoce como *ley de autoconciencia o introspección*.

Estas observaciones explican la íntima relación de la percepción con el pensamiento simbólico. La conducta educativa exige un proceso perceptivo más activo que uno analítico o reproductivo; supone problematizar lo familiar, así como desarrollar la comprensión y resolver problemas. Así definida, la capacidad educativa estaría relacionada con la *inteligencia fluida*, mientras que la capacidad reproductiva se aproxima a la *inteligencia cristalizada* (Raven, J. C., Raven, J. et al., 2003). La inteligencia fluida se refiere a la habilidad para utilizar el razonamiento abstracto para resolver nuevos problemas que no fueron enseñados en la escuela. Generalmente se evalúa mediante analogías abstractas cuyos datos no envejecen a lo largo del tiempo ni son específicos de una cultura. La inteligencia cristalizada, en cambio, es culturalmente más específica, dependiente de la escolarización formal y sujeta a la variedad de los conocimientos adquiridos. Por ende, el incremento de la inteligencia cristalizada a lo largo de la vida de una persona es esperable, mientras que el incremento de la inteligencia fluida no lo es (Kaufman, 2009).

Las Matrices miden la capacidad para educir relaciones. Esto se debe a que las variables entre las que deben verse relaciones no son en sí

mismas obvias. Hay que discernir la relación para reconocer esas variables. Por lo tanto, puede sostenerse que este test mide la inteligencia fluida. Coincidimos con Fernández Liporace, Ongarato, Saavedra y Casullo (2004, p.52) al afirmar que “esto explica la utilización tan difundida de las Matrices en los ámbitos de investigación” ya que los resultados que aporta permiten comparar poblaciones y/o sujetos que “han sido expuestos a situaciones educativas formales e informales de naturaleza bien disímil”.

En tanto prueba de evaluación psicológica, el Test de Raven integra el conjunto de instrumentos que permiten al profesional psicólogo arribar a un diagnóstico respecto de aquellos sujetos sobre los que recae la evaluación. Estos aportan información crucial para arribar a un pronóstico y planificar el desarrollo de estrategias de intervención específicas para cada caso. Las puntuaciones directas que obtienen los sujetos son transformadas a escalas que permiten su comparación con una muestra estandarizada. Los baremos sistematizan las informaciones otorgan significación estadística al desempeño individual. Pero para que la interpretación diagnóstica sea correcta los baremos deben estar actualizados, es decir, el grupo de referencia del cual se obtienen las puntuaciones promedio con las que se compara el rendimiento de un sujeto, debe ser el adecuado.

Casullo (2009:28) resalta la necesidad de plantearnos universales al proponernos estudiar una porción de la conducta a través de una técnica de evaluación psicológica; pero al momento de construirla, esta debe respetar la particularidad de la cultura “donde va a ser aplicada”. De esta manera, la revisión periódica de los tests en función de los nuevos descubrimientos e hipótesis que aportan las teorías así como los cambios culturales y las adaptaciones regionales, se vuelve necesaria para que el test aplicado sea válido y confiable. Siguiendo a Martorell Pallas y Gómez Llorens (2010, p. 39), la evaluación psicológica “no solo debe llevar a la descripción, diagnóstico o predicción sino que también debe incluir la valoración de la predicción y la valoración de los resultados obtenidos”.

La prueba fue revisada en sucesivas oportunidades, generalmente con el fin de obtener

normas o baremos adecuados a diferentes poblaciones.

El Efecto Flynn

A partir del uso generalizado de los tests de inteligencia, se ha observado que los puntajes medios obtenidos aumentan regularmente y de manera notable con el correr del tiempo en todo el mundo. James R. Flynn sintetiza los resultados de las investigaciones sobre el tema en el año 1984 y, a raíz de ello, Herrnstein y Murray (1994) denominaron “efecto Flynn” (Flynn, 1984) al aumento constante de los valores absolutos en los tests de inteligencia, para todas las edades, aludiendo a que, aunque se trataba de un fenómeno ya señalado, Flynn jugó un papel fundamental para que se le prestara la debida atención¹.

Esta constatación indica que, para la misma población, las normas se vuelven obsoletas conforme pasan los años. Es decir que en la actualidad, para obtener el mismo puntaje transformado que hace algunas décadas, las puntuaciones directas deberían ser más altas y para ello es necesario resolver un mayor número de problemas.

Estos aumentos son mayores en aquellos tests que miden la inteligencia fluida, que en los que miden la inteligencia cristalizada (Sundet, Barlaug, & Torjussen, 2004). Esto llevó a pensar que el aumento está relacionado con el factor G de la inteligencia y no con factores específicos (i.e., factores E).

La causa de este fenómeno permanece desconocida y no existe evidencia suficiente que permita afirmar que refleje un verdadero aumento de la inteligencia. Generalmente se hipotetiza que esté ligado a factores tales como la heterosis; la mejora en las condiciones de vida en las distintas poblaciones, entre las que se encuentra la alimentación o nutrición más adecuada, la reducción del tamaño de las familias nucleares, la expansión del sistema educativo con mayor desarrollo de la escolarización, con el consecuente incremento del desarrollo psicológico infantil; la

¹ “We call it the Flynn effect because of psychologist Flynn’s pivotal role in focusing attention on it, but the phenomenon itself was identified in the 1930s when testers began notice that IQ scores often rose with very successive year after a test was first standardized.” (Herrnstein & Murray, 1996, p. 307).

adquisición progresiva de ciertas habilidades para responder satisfactoriamente a los tests; la creciente tecnologización de la cultura, desde los juegos de video hasta el acceso cada vez más generalizado a los medios de comunicación a través de internet, que ofrecen otros aspectos de estimulación (Sundet et al., 2004, Prensky, 2001).

Por su parte, Armstrong y Woodley (2014) afirman que hay evidencia neurológica que señala que el *Efecto Flynn* está asociado a un aumento en el tamaño del cerebro que ha potenciado las funciones del hipocampo, aunque no descartan que otra causa para este fenómeno sea la posibilidad de que los tests que miden la inteligencia fluida y la capacidad eductiva (tal como lo hace el Test de Matrices Progresivas de Raven), generen algún tipo de “andamio cognitivo” que permita el uso de habilidades que no dependan del factor G para resolver los problemas.

En su estudio inicial, Flynn (1984) estableció que la magnitud del aumento del cociente intelectual –medido inicialmente con las escalas Wechsler y Stanford-Binet– era de 0.3 puntos por año o 3 puntos por década. Estos estudios se realizaron comparando muestras estadounidenses entre los años 1932 y 1978. Más adelante, la misma tendencia se corroboró con datos de otros 20 países.

No obstante, investigaciones recientes muestran un *efecto contrario*: una disminución en el crecimiento de dichos puntajes, observado por primera vez en los países escandinavos. Tal es el caso de investigaciones longitudinales hechas en Noruega (Sundet, Barlaug, & Torjussen, 2004) en las cuales se muestra desde los años 50 una disminución en los puntajes de una batería de tests administrados a los jóvenes de 18 años, antes de entrar al servicio militar. Las pruebas utilizadas son tests de matemática y lenguaje, similares a los subtests del WAIS IV, y un test no verbal que fue construido de forma similar al test de Raven. Los dos primeros miden la inteligencia cristalizada, mientras que el último mide la inteligencia fluida.

Teasdale y Owen (2007) estudiaron los datos aportados por una batería similar en Dinamarca, que se administra desde 1957 a todos los jóvenes de 18 años que ingresan al servicio militar. De los cuatro tests que la componen, los autores afirman

que el de Matrices de Letras se asemeja a las Matrices Progresivas de Raven. Entre sus resultados señalan, en jóvenes de todos los niveles educativos, que si bien hubo un pequeño incremento de los puntajes entre los años 1988 y 1998, los mismos disminuyeron en el período 2003-2004, incluso por debajo de los obtenidos en 1988 para el conjunto de los tests de la batería. Esos resultados apoyan la afirmación de que en lo que va del siglo XXI ha habido muy poca evidencia sobre la continuidad del efecto Flynn, tal como se lo ha definido originalmente, es decir, como el aumento *continuo* de las medias en los tests de inteligencia.

Por su parte, Brouwers, Van de Vijver y Van Hemert (2008) hicieron una revisión de 193 estudios sobre las tres escalas Raven publicados entre 1944 y 2003, que abarcan 798 muestras provenientes de 45 países, con un número 244.316 sujetos en total. Dentro de ese reporte, Argentina se encuentra representada por más de 20 estudios publicados en dicho período. Los autores calcularon la correlación entre las características de las muestras (edad, años de escolaridad), de los países de los que provenía cada estudio (Producto Bruto Interno, escolaridad y analfabetismo), y de los estudios consultados (año de publicación). Encontraron que, si bien tanto el PBI como los años de escolaridad se correlacionan positivamente con el desempeño en el test (por lo que pueden estar influyendo en las diferencias interculturales), existe una correlación negativa significativa entre el PBI y la magnitud del Efecto Flynn. A partir de esto, concluyen que habría alcanzado su techo en los países desarrollados de occidente, mientras que los países con un PBI per cápita menor, todavía muestran valores crecientes. Esta conclusión coincide con la de Teasdale y Owen (2007), quienes afirman que si el efecto Flynn llega a su fin en los países más desarrollados, estaría lejos de hacerlo en el resto del mundo, aunque se esperaría que las diferencias entre los distintos países en el futuro disminuyeran gradualmente.

El mismo Flynn (2013) también ha subrayado que en Holanda y Finlandia aparece una disminución en la magnitud del efecto que lleva su nombre, pero que sin embargo, esta tendencia no se evidencia en todos los países desarrollados. Los

datos recientes provenientes de Australia son ambiguos al respecto y Estados Unidos y Corea del Sur constituyen claras excepciones a la misma. También menciona que algunos autores atribuyeron esta tendencia a la inmigración a los países desarrollados de poblaciones con CI bajos, no obstante esto no explicaría totalmente la situación de Estados Unidos. Flynn aventura la idea de que la detención de los incrementos del CI en ciertos países podría deberse a que en ellos ya no puede progresarse más en aquellos aspectos que se supone los fomentaría.

A su vez, Rusell (2007) sostiene que un efecto meseta sería de esperar en culturas que tengan condiciones de vida óptimas, lo cual explicaría la razón por la cual se ha manifestado en los países escandinavos antes que en el resto del mundo. Efectivamente en ellos, desde el fin de la Segunda Guerra Mundial, se ha producido un sistema de bienestar que incluye a toda la población. Recurriendo a las escalas Wechsler y con el procedimiento de análisis de regresión lineal, este autor prevé que en Estados Unidos el efecto meseta se alcanzará en 2024. Pero al considerar posibles sesgos en la selección de las muestras para las estandarizaciones previas de estas escalas, estima que la meseta podría haberse alcanzado más tempranamente en el año 2004, por lo que el efecto Flynn podría ya no estar vigente en ese país.

Con respecto a la Argentina, este equipo realizó una primera constatación del Efecto Flynn en la ciudad de La Plata y alrededores al actualizar los baremos del Test de Matrices Progresivas de Raven en el año 2000, y compararlos con la estandarización previa del año 1964. Esta comparación mostró un considerable aumento en los puntajes directos del test para todas las edades (Flynn & Rossi-Casé, 2012; Rossi-Casé, Neer, & Lopetegui, 2001, 2002, 2011). Asimismo, se realizó un análisis de los resultados por género que sirvieron para comprobar tendencias similares a la encontradas a nivel internacional (Flynn & Rossi-Casé, 2011).

Por todo lo señalado hasta aquí puede advertirse que las normas de los tests psicológicos en modo alguno son absolutas, universales o permanentes (Anastasi, 1980). En este marco, y dada la importancia del Test de Matrices

Progresivas de Raven como instrumento de evaluación psicológica y su uso generalizado con diferentes propósitos en distintos ámbitos de aplicación de la psicología, se hace necesaria la actualización de los baremos para poder utilizarlo de forma confiable. Tal es el objetivo fundamental de este trabajo, así como comparar los resultados con las normas obtenidas en 1964 y 2000, para muestras de iguales características. Asimismo nos proponemos introducir un análisis diferenciado individualmente de las series que componen el test, para examinar el orden de dificultad de los ítems que la componen.

Método

Participantes

La muestra de estandarización consistió en 1050 sujetos de ambos sexos, con edades que variaban de los 13 a los 18 años, cumplidos al momento de la administración del test. En el presente estudio se dividió el rango de edad de 13 a 18 años, en tres intervalos iguales, teniendo en cuenta que el desempeño en la prueba puede cambiar de modo atendible en cada uno ellos, razón por la cual se justifica la generación de normas diferentes para cada grupo de edades.

La muestra fue estratificada con respecto a edad, género y tipo de gestión de los colegios (de gestión pública o privada), pertenecientes a la educación secundaria y ubicados en la ciudad de La Plata. Se extrajeron datos del censo poblacional del año 2010 (Instituto Nacional de Estadística y Censos [INDEC], 2010). Se utilizó un muestreo bietápico que permitió seleccionar distintos conglomerados (instituciones educativas) y en cada uno de ellos elegir aleatoriamente los grupos de clase que serían examinados para las distintas edades, siendo evaluados todos los sujetos que conformaban cada grupo. Se trabajó con un intervalo de confianza del 95% que permitió establecer que el error muestral máximo sería del 3% para este tamaño de muestra. En la tabla 1, se describe la composición de la muestra, según sexo y edad. En la tabla 2 se comparan los porcentajes de la composición poblacional y muestral, por rango de edades.

Tabla 1. Composición de la muestra de tipificación del Test de Raven, año 2014, La Plata (Argentina)

Edad	Sexo		Total de casos
	Mujer	Hombre	
13 años	73	135	208
14 años	68	85	153
Totales por grupo de 13-14 años	141	220	361
15 años	82	99	181
16 años	83	72	155
Totales por grupo de 15-16 años	165	171	336
17 años	90	102	192
18 años	52	101	160
Totales por grupo de 17-18 años	130	174	352
Total de casos por sexo	455	594	1049

Nota. Fuente: Muestra, Elaboración propia basada en la administración del Test de Raven-Escala General, en 2012-2014, a estudiantes secundarios de ambos géneros, en La Plata, Argentina ($N=1049$). Test de Raven, Carpeta de Evaluación, 2005, Ed. Paidós. Bs. As.

Instrumentos

Se utilizó el Test de Matrices Progresivas de Raven; Escala General, segunda edición (Raven, Raven, & Court, 2003). El test está compuesto por 60 problemas organizados en cinco series (A, B, C, D y E), de doce ítems cada uno. En cada una de las series los ítems están ordenados por dificultad creciente, comenzando con problemas sencillos. Cada serie resulta con un grado de dificultad mayor que la anterior, pero de modo tal que el último problema de una serie resulta más difícil que el primero de la serie inmediata posterior.

Procedimientos

El test se administró de manera colectiva, en grupos-clase, con presencia del examinador y sin límite de tiempo para su ejecución, para permitir la evaluación de la capacidad intelectual sin la intervención de la velocidad para resolver la tarea. Para la participación de los sujetos de la muestra se solicitó, en tiempo y forma, la debida autorización a los Inspectores Educativos del Distrito La Plata y se obtuvo el consentimiento informado previo de los padres o tutores de los adolescentes, mediante un formulario donde se explicaba el propósito de la investigación y la

confidencialidad de los datos. El test fue administrado en distintos períodos de los años 2012, 2013 y 2014 (noviembre de 2012; junio, octubre y noviembre de 2013; mayo y junio de 2014).

Tabla 2. Comparación porcentual por rango de edades de la población y la muestra de tipificación del Test de Raven, La Plata (Argentina)

Grupo de edades	Censo poblacional 2010	Muestra 2014
13-14 años	32.44	34.41
15-16 años	32.75	32.03
17-18 años	34.81	33.56
Porcentaje total	100.00	100.00

Nota. Fuente: Población, Censo Nacional de Población 2010. Ministerio del Interior de la República Argentina. Muestra, Elaboración propia basada en la administración del Test de Raven-Escala General, en 2012-2014, a estudiantes secundarios de ambos géneros, en La Plata, Argentina ($N=1049$). Test de Raven, Carpeta de Evaluación, 2005, Ed. Paidós. Bs. As.

La administración estuvo a cargo de profesionales y alumnos avanzados de la carrera de Psicología, entrenados previamente a fin de homogeneizar el procedimiento. Para la consigna, se siguieron de manera rigurosa las instrucciones dadas por el autor del test para la administración colectiva. La administración se llevó a cabo en cada una de las escuelas seleccionadas, en horarios regulares de clases, cedidos por las autoridades y profesores de cada institución. Cada grupo no superó los 30 alumnos.

Al conformar la base de datos, se excluyeron aquellos protocolos que mostraban una discrepancia mayor a la esperada en la composición de los puntajes ($n = 146$). Finalmente se analizó una base de datos que incluye 1049 protocolos.

Análisis

Con la información obtenida se conformó una base de datos (Programa Excel, Microsoft, 2007) obteniéndose los estadísticos descriptivos correspondientes así como los percentiles que permitieron elaborar las normas para este grupo.

Los resultados obtenidos en el estudio actual se compararon con las normas obtenidas en los años 1964 y 2000 para los mismos grupos etarios de la ciudad de La Plata, Argentina, utilizando el estadístico t de Student con un nivel de confianza de 95%. El análisis de datos se realizó por grupos de edad: 13-14 años; 15-16 años y 16-17 años, para poder comparar los valores percentilares con otros estudios de la misma índole. Para los mismos grupos se compararon además los puntajes obtenidos en las series para poder estimar su grado de dificultad.

Resultados

La descripción de las puntuaciones directas obtenidas muestra que en los tres grupos de edad el rango de respuestas correctas varía de 10 a 60 (en este último caso la puntuación máxima posible). Las puntuaciones mínimas resultaron similares para todas las edades. Los resultados, expuestos en la tabla 3, mostraron que a medida que aumenta la edad de los sujetos, el rendimiento de los grupos mejora y la heterogeneidad de sus respuestas disminuye.

Tabla 3. Test de Raven. Puntuaciones promedio, desviación estándar y rango según edades, año 2014, La Plata (Argentina)

Grupo de Edades	Promedio	Desviación estándar	Rango de puntajes
13-14 años	44.01	7.60	12 a 60
15-16 años	48.39	6.75	10 a 60
17-18 años	49,23	6,39	12 a 60

Nota: $N = 1049$. Puntuación máxima: 60 puntos.

La distribución de las puntuaciones directas necesarias para la ubicación de los sujetos en la escala percentilar se presenta en la Tabla 4.

El análisis de las puntuaciones promedios de cada Serie, por rango de edades, muestra que el orden creciente de dificultad para la resolución de los ítems del Test se altera en el caso de las series C y D. En todos los grupos de edad estudiados los puntajes de la primera resultaron menores a los de la segunda (Ver Tabla 5).

Tabla 4. Test de Raven, Normas 2014, Escala General. La Plata (Argentina)

Percentil	Puntuación directa según edades		
	13-14 años	15-16 años	17-18 años
P99	57	58	59
P95	54	56	57
P90	52	55	56
P75	50	53	54
P50	45	50	50
P25	40	46	46
P10	34	40	40
P5	30	35	38

Nota: Fuente: Muestra, Elaboración propia basada en la administración del Test de Raven-Escala General, en 2012-2014, a estudiantes secundarios de ambos géneros, en La Plata, Argentina ($N=1049$). Test de Raven, Carpeta de Evaluación, 2005, Ed. Paidós. Bs. As.

Tabla 5. Promedio de las puntuaciones directas en cada serie del Test de Raven, año 2014, La Plata (Argentina)

Grupo de Edades	Serie A	Serie B	Serie C	Serie D	Serie E
13-14 años	11,30	10,39	8,57	9,02	4,71
15-16 años	11,57	11,01	9,63	9,88	6,31
17-18 años	11,65	10,93	9,76	10,22	6,66

Nota: ($N=1049$). Test de Raven, Carpeta de Evaluación, 2005, Ed. Paidós. Bs. As. Puntuación máxima por Serie: 12 puntos

Resultados para el grupo de edad 13-14 años

En los sujetos que integran este grupo se observan los resultados más bajos de todos los que integran la muestra. La puntuación promedio es de 44.05 puntos y la desviación estándar es de 7.60 puntos. El rango de respuestas correctas fue de 12 a 60. La mitad de los sujetos obtuvo 45 puntos o menos. Si bien esto representa un incremento de 6 puntos respecto del baremo del año 1964, el resultado observado actualmente es menor en 3 puntos que el alcanzado en el año 2000. Este incremento significativo de las puntuaciones respecto de las normas 1964 ($t(359) = 1.98$; $p < .05$) y la disminución respecto de los baremos del año 2000, se observó en todos los valores percentilares calculados. Los resultados se muestran en la Tabla 6.

Tabla 6. Test de Raven. Comparación de las Puntuaciones Directas de sujetos de 13-14 años en los años 1964, 2000 y 2014, La Plata (Argentina)

Percentil	Puntuaciones Directas		
	año 1964	año 2000	año 2014
P99	50	59	57
P95	49	56	54
P90	48	55	52
P75	44	52	50
P50	39	48	45
P25	35	44	40
P10	27	39	34
P5	22	36	30

Nota: Elaboración propia basada en la administración del Test de Raven-Escala General (Raven, J. C., Raven, J. et al., 2003), en 2012-2014, a estudiantes secundarios de ambos géneros, en La Plata, Argentina ($n = 361$).

Resultados para el grupo de edad 15-16 años

La puntuación promedio observada es de 48.39 puntos y la desviación estándar es de 6.75 puntos. El rango de respuestas correctas fue de 10 a 60.

La mitad de los sujetos obtuvo al menos 50 puntos. Esto representa un incremento de 9 puntos respecto del baremo del año 1964, y de 1 punto respecto de la norma alcanzada en el año 2000. Este incremento significativo de las puntuaciones respecto de las normas 1964 se observó en todos los valores percentilares calculados ($t(334) = 1.41$; $p < .05$). La comparación de resultados con los baremos del año 2000 muestra aumento de 1 y 2 puntos para los percentiles 25, 50 y 75, mantenimiento de los puntajes para los percentiles 10, 90 y 95; y una disminución de 1 punto para percentiles 5 y 99. Los resultados se muestran en la Tabla 7.

Tabla 7. Test de Raven. Comparación de las Puntuaciones Directas de sujetos de 15-16 años, Baremos 1964, 2000 y 2014, La Plata (Argentina)

Percentil	Puntuaciones directas		
	año 1964	año 2000	año 2014
P99	52	59	58
P95	51	56	56
P90	49	55	55
P75	44	52	53
P50	41	49	50
P25	36	44	46
P10	28	40	40
P5	23	36	35

Nota: Elaboración propia basada en la administración del Test de Raven-Escala General (Raven, Raven, & Court, 2003), en 2012-2014, a estudiantes secundarios de ambos géneros, en La Plata, Argentina ($n=336$)

Resultados para el grupo de edad 17-18 años

En los sujetos que integran este grupo se observan los resultados más altos de todos los que integran la muestra. La puntuación promedio es de 49.23 puntos y la desviación estándar de 6.39 puntos. El rango de respuestas correctas fue de 12 a 60.

La mitad de los sujetos obtuvo 50 puntos o menos. El resultado observado representa un incremento de 10 puntos respecto del baremo del año 1964, al tiempo que permanece igual a la norma alcanzada en el año 2000. Este incremento significativo de las puntuaciones respecto de las

normas 1964 ($t(302) = 0.64$; $p < .05$) se observó en todos los valores percentilares calculados. Respecto de los baremos del año 2000, se constata la disminución de 3 y 2 puntos para los valores percentiles 5 y 10; igualdad en las puntuaciones de los percentiles 25, 50, 90 y 95; y un aumento de un 1 punto en los valores de los percentiles 75 y 99. Los resultados se muestran en la Tabla 8.

Tabla 8. Test de Raven. Comparación de las Puntuaciones Directas del grupo 17-18 años, años, Baremos 1964, 2000 y 2014, La Plata (Argentina)

Percentil	Puntuaciones Directas		
	año 1964	año 2000	año 2014
P99	53	58	59
P95	53	57	58
P90	50	56	56
P75	45	53	54
P50	40	50	50
P25	35	46	46
P10	27	43	41
P5	22	40	38

Nota: Elaboración propia basada en la administración del Test de Raven-Escala General (Raven, Raven, & Court, 2003), en 2012-2014, a estudiantes secundarios de ambos géneros, en La Plata, Argentina ($n = 352$).

Resultados comparativos para el grupo de edad 13-16 años y 17-18 años

La metodología aplicada para la descripción de los resultados obtenidos en el baremo del año 2000 permite mostrar que el rendimiento promedio de los sujetos de 13 a 16 años en el estudio actual es 1.64 puntos menor que aquel; y la variabilidad de las respuestas es 1.33 puntos mayor en la actualidad. Para los sujetos del grupo de edad 17-18 años, los rendimientos promedio

son levemente inferiores, y la variabilidad aumenta en 1.13 puntos. Los resultados se muestran en la Tabla 9.

Discusión y Conclusiones

Respecto de las normas del año 1964 y las correspondientes al año 2000, los resultados observados corroboran el efecto Flynn. Sin embargo, entre el baremo del año 2000 y el estudio actual, dicho efecto desaparece. En consonancia con investigaciones recientes realizadas en otros países (Brouwers, Van de Vijver y Van Hemert, 2008; Rusell, 2007; Sundet et al., 2004; Teasdale y Owen, 2007), los resultados obtenidos permiten mostrar una detención del aumento de los puntajes directos necesarios para alcanzar el rendimiento promedio en cada rango de edad. En el estudio actual se constata una leve disminución de las puntuaciones promedio que no resulta significativa, por lo que hablamos de un amesetamiento de los puntajes.

Si el efecto Flynn evidencia que los puntajes han ido creciendo a razón de más o menos una desviación estándar por generación (Flynn, 1984), resulta adecuado preguntarse si este aumento será indefinido a lo largo del tiempo o por el contrario en algún momento cesará y en tal caso cuando. Para intentar dar respuesta a esta cuestión deberíamos volver a analizar las hipótesis que en su momento intentaron explicar la presencia del efecto Flynn, tales como mejoras en la alimentación, mayor escolarización, el uso de las nuevas tecnologías y el predominio de las imágenes visuales, entre otras.

Tabla 9. Test de Raven. Comparación de las puntuaciones directas y desvíos estándar, por grupos de edades, años 2000 y 2014, La Plata (Argentina)

Grupos de edad	Puntuaciones Directas			
	año 2000		año 2014	
	Promedio	Desviación estándar	Promedio	Desviación estándar
13-16 años	47.76	6.19	46.12	7.52
17-18 años	49.50	5.44	49.23	6.39

Nota: Elaboración propia basada en la administración del Test de Raven-Escala General (Raven, Raven, & Court, 2003), en 2012-2014, a estudiantes secundarios de ambos géneros, en La Plata, Argentina (N=1049).

Observamos que la presencia del estímulo visual se tornó común y generalizada en este período, y que los participantes de este estudio, son en su mayoría nativos tecnológicos de la aldea global (Prensky, 2001). Al mismo tiempo, la facilidad de acceso a la información generaliza la difusión de los conocimientos y avances relacionados con la salud y la nutrición (Rossi-Casé, Neer y Lopetegui, 2001a).

Entonces, si las condiciones que alguna vez se mencionaron para explicar el efecto Flynn se han estabilizado en los últimos años, los puntajes, que eran su expresión, han corrido la misma suerte. No habiendo ninguna otra condición que produzca un cambio cualitativo del mismo tenor, las puntuaciones directas se mantienen estables, lo que condice con el amesetamiento observado en Argentina y otros países del mundo.

Otra posible explicación de este fenómeno podría ser que el test de Matrices Progresivas de Raven no estuviese midiendo únicamente lo que dice medir, y que otros factores fueran los que incidan en su resolución, como por ejemplo la educación formal y, en consecuencia, la inteligencia cristalizada. Estos factores podrían ser más sensibles a los estímulos del ambiente, lo que podría explicar no solo los cambios que aparecen a lo largo de los años, sino también las diferencias que existen entre los países desarrollados y aquellos en vías de desarrollo.

Más allá del fenómeno estudiado, un aspecto llamativo que surge del análisis de los resultados es que no se cumple con la progresión de dificultad en la resolución de las matrices de las series C y D. Para los tres grupos de edad estudiados, el promedio de las puntuaciones directas de la serie C mostró ser más bajo que las de la serie D, resultado contrario a lo esperado según el diseño del Test de Raven. Los mismos coinciden con los observados en investigaciones realizadas por Cairo-Valcárcel (2000) y Fernández-Liporace (2004). En ambos casos se sugiere la necesidad de examinar con mayor profundidad estos resultados con el propósito de reestructurar las series C y D para devolver a la prueba su original diseño de dificultad creciente. Esta graduación puede incidir sobre los sujetos examinados, en la medida en que deben esperar mayor dificultad cuando las matrices se van

sucedendo y no en sentido contrario, tal como se observa en el presente estudio.

Finalmente, las conclusiones principales a las que se arriba pueden resumirse como la necesidad de actualizar los baremos, así como la correspondiente al análisis de los instrumentos, a la luz de los resultados encontrados, de modo tal de contar con datos precisos a la hora de su aplicación con fines investigativos, educacionales o clínicos.

Referencias

- Anastasi, A. (1980). *Tests Psicológicos*. Madrid: Aguilar.
- Armstrong, E. L., & Woodley, M. A. (2014). The rule-dependence model explains the commonalities between the Flynn effect and IQ gains via retesting. *Learning and Individual Differences*, 29, 41-49. Recuperado de <http://dx.doi.org/10.1016/j.lindif.2013.10.009>
- Brouwers, S. A., Van de Vijver, F. J. R., & Van Hemert, D. A. (2008). Variation in Raven's Progressive Matrices scores across time and place. *Learning and Individual Differences*, 19, 330-338. Recuperado de <http://dx.doi.org/10.1016/j.lindif.2008.10.006>
- Cairo-Valcárcel, E., Cairo-Martínez, E., Bouza, C., & Ponce-Zolozaval, T. (2000). Algunas características y posibilidades del Test de Matrices Progresivas de Raven. *Revista Cubana de Psicología*, 17(2), 95-105. Recuperado de <http://pepsic.bvsalud.org/pdf/rcp/v17n2/02.pdf>
- Casullo, M. M. (2009). La evaluación psicológica: modelos, técnicas y contextos. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica*, 27(1), 9-28.
- Cattel, R. B. (1968). The theory of fluid and crystallized intelligence. Its relationship to culture free tests and its verification in 9-12 ys. old children. *Bolletino di Psicologia Applicato*, 88(90), 3-22.
- Fernández-Liporace, M., Ongarato, P., Saavedra, E., & Casullo, M. M. (2004). El Test de Matrices Progresivas, Escala General: Un análisis psicométrico. *Evaluar*, 4, 50-69.

- Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932-1978. *Psychological Bulletin*, 95, 29-51.
- Flynn, J. R. (2013). The "Flynn Effect" and Flynn's paradox. *Intelligence*, 41(6), 851-857. doi:10.1016/j.intell.2013.06.014
- Flynn, J. R., & Rossi-Casé, L. (2011). Modern women match men on Raven's Progressive Matrices. *Personality and Individual Differences*, 50(6), 799-803. <http://dx.doi.org/10.1016/j.paid.2010.12.035>
- Flynn, J. R., & Rossi-Casé, L. (2012). IQ gains in Argentina between 1964 and 1998. *Intelligence*, 40, 145-150. Recuperado de <http://dx.doi.org/10.1016/j.intell.2012.01.006>
- Herrnstein, R. J., & Murray, Ch. (1996). *The Bell Curve. Intelligence and Class Structure in American Life*. New York: Free Press Paperbacks. Simon & Schuster
- Instituto Nacional de Estadística y Censos. (2010). *Censo Nacional de Población, Hogares y Viviendas 2010*. Recuperado de <http://www.censo2010.indec.gov.ar/>
- Kaufman, A. S. (2009). Hot Topic: Are our IQs fixed or are they malleable? En J. C. Kaufman (Ed.), *IQ Testing 101*, (pp. 201-222). New York: Springer.
- Martorell Pallas, C., & Gómez Llorens, O. (2010) Enfoque de la Evaluación Psicológica de la Revista Iberoamericana de Diagnóstico y Evaluación Psicológica (Ridep). *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 30(2), 35-56.
- Prensky, M. (2001). *Digital natives, digital immigrants. On the Horizon*. Recuperado de https://bemidjistate.ims.mnscu.edu/d2l/lms/content/viewer/main_frame.d2l?ou=1719951&tid=13826298
- Raven J. C., Raven, J., & Court, J. H. (2003). *Test de Matrices Progresivas: Escala General*. Buenos Aires: Paidós.
- Rossi-Casé, L., Neer, R., & Lopetegui, S. (2001). Test de Matrices Progresivas de Raven: Comparación de baremos. El aumento de los puntajes directos a través del tiempo. *Evaluar*, 2(2), 39-51.
- Rossi-Casé, L., Neer, R., & Lopetegui, S. (2002). Test de Matrices Progresivas de Raven: Construcción de Baremos y Constatación del "Efecto Flynn". *Orientación y Sociedad*, 3, 181-187.
- Rossi-Casé, L., Neer, R., & Lopetegui, S. (2011). Baremo ciudad de La Plata (2000). Buenos Aires, Argentina. Escala General (MPG). En J. C. Raven (Ed.), *Test de Matrices Progresivas. Carpeta de Evaluación, Escala General* (pp. 19-23). Buenos Aires: Paidós.
- Rusell, E. W. (2007). Commentary: The Flynn effect revisited. *Applied Neuropsychology*, 14(4), 262-266. Recuperado de <http://dx.doi.org/10.1080/09084280701719211>
- Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32, 349-362. <http://dx.doi.org/10.1016/j.intell.2004.06.004>
- Teasdale, T. W., & Owen, D. R. (2007). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence*, 36, 121-126. Recuperado de <http://dx.doi.org/10.1016/j.intell.2007.01.007>