

# Una Revisión Crítica de la Escala Terman. Por Qué No el Stanford-Binet Tercera Edición Forma L-M

## A Critical Review of the Terman Scale. Why we Should Not Use the Third Edition Stanford-Binet Form L-M

Juan José Ballesteros Domingo  
EOEP Centro-Arganzuela MADRID

**Resumen.** La escala Terman (Stanford-Binet Terman-Merrill Tercera Edición, conocida también como Forma L-M) fue utilizada de manera general en España hasta finales de los años 70. Actualmente su uso es minoritario y restringido casi exclusivamente para la evaluación y diagnóstico de escolares con posible alta capacidad intelectual. En los últimos años se va haciendo más frecuente la demanda (en algunos casos, exigencia) por parte de los padres de que se utilice o, al menos, se consideren los resultados de la escala. La finalidad de este artículo es exponer, a partir de las características principales de la escala, los argumentos fundamentales (todos suficientes) para no utilizarla en la actualidad ni considerar los resultados obtenidos con ella: (1) ausencia de normas válidas, (2) heterocedasticidad, (3) efectos del cambio en la naturaleza del CI, (4) dificultades de integración en el contexto de la evaluación psicopedagógica y (5) dificultades de asimilación del concepto de edad mental para los destinatarios.

*Palabras clave:* medida de la inteligencia, normas y heterocedasticidad de escala Terman.

**Abstract.** The Terman scale (Stanford-Binet Terman-Merrill Third Edition, also known as Form L-M) was generally used in Spain until the late 70s. Presently its use is marginal and restricted almost exclusively to the evaluation and diagnosis of students with a possible high intellectual capacity. During the last few years parents have increasingly requested (and sometimes demanded) the use, or at least the consideration of the results of the scale. The purpose of this article is to present, from the main characteristics of the scale, the fundamental arguments against its use at present and in favour of disregarding its results: (1) absence of valid norms, (2) heterocedasticity, (3) effects of the change in the nature of IQ, (4) difficulties of integration within the context of psychopedagogical evaluation, and (5) difficulties assimilating the concept of mental age for the recipients.

*Key Words:* Measurement of Intelligence, norms and heterocedasticity of the Terman scale.

El máximo desarrollo de las escalas de edad tipo Binet se alcanzó con la tercera edición de la Stanford-Binet, realizada por M. A. Merrill en 1960<sup>1</sup>. Fue de uso general en nuestro país, conocida en términos coloquiales, igual que las otras versio-

nes, como *el Terman*, hasta que se impusieron las de estructura factorial tipo Weschler. Aunque existen numerosas referencias y apelaciones al uso de la escala, prácticamente no se encuentra información sobre la misma, aparte de la ofrecida por los propios

---

La correspondencia sobre este artículo debe enviarse al autor al E-mail: [juanjoseballes@yahoo.es](mailto:juanjoseballes@yahoo.es)

---

<sup>1</sup> L. M. Terman falleció en diciembre de 1956, aunque sigue apareciendo como autor en la Tercera Edición y sus referencias.

manuales. La escala Terman se reivindica como válida (especialmente para la evaluación de las altas capacidades) pero apenas se fundamenta esta pretensión, ni siquiera por sus partidarios más convencidos. Los profesionales más jóvenes sólo la han estudiado como parte de la historia de la Psicología, sin haber tenido ocasión de manejarla. Para facilitar la exposición posterior, describo a continuación brevemente sus principales características, aunque considero necesario consultar con detalle, al menos, el último manual de la escala (Terman y Merrill, 1975).

### **Estructura, aplicación, corrección y normas de la escala Terman**

La tercera edición de la escala Terman es, como las anteriores, una escala tipo Binet, cuya finalidad es la de medir la inteligencia como capacidad general, más que como capacidades específicas o grupos de ellas en función de teorías que las sustenten. Se puede aplicar a sujetos a partir de los dos años de edad cronológica (EC). Emplea el concepto de Edad Mental (EM), siendo ésta la unidad, en meses, que se utiliza para medir el rendimiento del sujeto en la escala. Está compuesta por 142 pruebas breves, repartidas en grupos de seis, más una suplementaria a lo largo de veinte niveles, uno por año de edad mental (salvo entre los dos y los cinco años, que se escalonan cada seis meses y a partir de catorce años que se organizan en un nivel de Adultos Medios y tres niveles de Adultos Superiores). Del total, sólo 36 pruebas se pueden considerar no-verbales, la mayoría de ellas (24) para edades mentales menores de seis años. (Becker, K.A., 2003).

Las pruebas son superadas cuando el sujeto alcanza el *criterio de paso* para cada una de ellas, rendimiento mínimo suficiente para la tarea propuesta, siendo el resultado dicotómico (superada / no superada). Un reducido número de pruebas se valoran y puntúan en más de un nivel de edad mental, con criterios de paso progresivamente exigentes. Destaca entre éstas últimas la de *Vocabulario*, que se valora

para todos los años pares de edad mental a partir de los seis y en todos los niveles de Adultos. Más adelante me referiré detalladamente a esta prueba. El procedimiento de aplicación permite que el sujeto no tenga que abordar todas las pruebas, sino sólo un número sensiblemente menor. Se inicia la escala por las pruebas situadas en el mayor nivel de edad mental en el que se espera que el sujeto supere con acierto todas ellas. Este nivel debe estimarse en función de la edad y cualquier otra información disponible sobre su capacidad. Si efectivamente el sujeto tiene éxito en todas las pruebas de una edad mental, ésta se considera *edad base* y sobre ella se van sumando meses con la superación de las pruebas siguientes. En general, no se abordan pruebas de una edad hasta no haber aplicado las de la anterior. El proceso termina cuando el sujeto falla todas las correspondientes a un nivel de edad mental. Para obtener la puntuación total, a la edad base comprobada se suman dos meses por cada prueba posterior superada (un solo mes para edades menores de cinco años). La puntuación total del sujeto es la EM medida en meses. El cociente intelectual (CI) se obtiene de las normas correspondientes, que relacionan la EC en el momento de realizar la prueba con la EM obtenida.

A medida que aumenta la EC, la media y la dispersión de las EM va también aumentando. Para cada edad, la relación entre EM y CI es lineal y las funciones que generan las normas de la escala para cada EC, son del tipo:  $CI = a + (b \times EM)$ , siendo gráficamente  $a$  la altura y  $b$  la pendiente de la recta que representa cada EC. Para que los valores de CI sean comparables de unas edades a otras, las pendientes de las rectas que representan las EC van disminuyendo. La Tabla 1, de elaboración propia, presenta los valores  $a$  y  $b$  de las funciones que generan para cada edad las tablas de la Segunda Edición de la escala y las normas de 1960 y 1972 de la Tercera Edición. Las funciones obtenidas a partir de la Tabla 1 pueden incluso sustituir a las normas correspondientes para cualquier valor de EM, esté o no incluido en dichas normas. Para edades intermedias deben extrapolarse los resultados, siendo más rápido el uso de las normas.

Tabla 1. Constantes **a**, **b** y **m** para las funciones generadoras de las Tablas (1937) y Normas (1960 y 1972) de la Escala Terman

EC	Segunda Edición (1937)		Tercera Edición (1960)		Tercera Edición (1972)	
	b	a	b	a	b	m
2 años	4,167	-1,31	4,125	-11,42	4,122	-2,77
2 a 6 m	3,333	10,41	2,893	-2,13	2,948	-0,72
3 años	2,778	8,44	2,444	-2,18	2,440	-0,89
3 a 6 m	2,381	5,68	2,197	-5,46	2,204	-2,48
4 años	2,083	0,70	2,038	-8,75	2,018	-4,33
4 a 6 m	1,852	-4,24	1,915	-13,87	1,900	-7,30
5 años	1,667	-9,21	1,813	-14,78	1,749	-8,45
5 a 6 m <sup>a</sup>	1,515	-12,34	1,702	-16,18	1,637	-9,88
6 años	1,389	-11,14	1,547	-13,19	1,478	-8,92
7 años	1,190	-7,59	1,271	-6,32	1,205	-5,25
8 años	1,042	-3,82	1,060	0,21	0,990	0,21
9 años	0,926	-0,77	0,909	3,70	0,851	4,34
10 años	0,833	2,81	0,787	7,59	0,729	10,41
11 años	0,758	7,64	0,686	8,75	0,661	13,23
12 años	0,694	10,42	0,613	10,53	0,595	17,69
13 años	0,641	10,57	0,569	10,21	0,554	18,42
14 años	0,609	8,43	0,550	8,72	0,529	16,48
15 años	0,581	6,50	0,533	7,96	0,505	15,77
16 años	0,555	1,51	0,529	6,10	0,481	12,67
17 años	—	-2,88	0,546	3,55	0,484	7,34
18 años	—	-5,05	0,554	1,718	0,487	3,53

Nota. para la Segunda Edición a = 0.

<sup>a</sup> se incluye la edad de 5 a 6 m por su interés estadístico, aunque no tenga pruebas propias.

### Ausencia de normas válidas

La estructura actual de la escala, la Forma L-M, quedó determinada en la primera versión aparecida en 1960. En realidad, se utilizó como base la Forma L de la Segunda Edición, sobre la que se modificaron los criterios de paso y la colocación de unas pocas pruebas, sustituyéndose algunas otras por las de la Forma M. Los criterios de paso de las pruebas, en general, resultaron más exigentes que para la Segunda Edición. (Terman y Merrill, 1975, Apéndice B). En esta versión se establecieron nuevas normas para la obtención del CI, eliminándose con ellas lo que se llamó la *variabilidad atípica* de algunas edades. Se abandonó definitivamente el cálculo del CI como el cociente EM / EC expresado en porcentaje. Medir el rendimiento en unidades de EM siguió considerándose por sus autores, como en ediciones anteriores de la escala, la manera menos artificiosa y más comprensible para los habituales destinatarios de sus resultados (padres,

maestros, profesores, trabajadores sociales, médicos, etc.), aunque ya no coincidieran los valores de EC y EM para el CI de 100.

En 1972 aparece la última actualización de normas, por considerarse que las anteriores de 1960 ya no representaban a la población norteamericana de esos años. Los criterios de paso y la distribución de las pruebas se mantuvieron sin cambios (salvo un estímulo visual en una de las pruebas de dos años y la sustitución de un término por otro en una de las pruebas de los siete años). (Terman y Merrill, 1975). Los resultados obtenidos por los sujetos de las muestras fueron superiores a los de 1960. Esto fue así en todos los niveles de edad pero especialmente en los menores de 6 años, que superaron de media el 110 de CI antes del ajuste.

El empleo de la escala Terman, aquí y ahora, sólo puede basarse en el supuesto de que las normas son universales e inmutables. Entre las revisiones de 1960 y 1972 hay diferencias significati-

vas en los resultados finales de CI. Las funciones de las normas de 1972 dan sistemáticamente menores valores de CI que las de 1960, para iguales rendimientos en EM. Conviene insistir en que la prueba sobre la que se aplicaron ambas revisiones era y sigue siendo hoy la misma. En una década el valor medio de rendimiento (EM) aumentó para todas las edades. Pasados los años 70, al no actualizarse más la escala ni haberse publicado nuevas normas, se desconoce la conversión actual en CI de las edades mentales para la población norteamericana. Nunca se han elaborado normas para la población española.

### Heterocedasticidad

El error estadístico (no me referiré aquí a otras fuentes de error) está determinado fundamentalmente por la variabilidad de las puntuaciones, que en las escalas de edad aumenta con el rendimiento (EM) más que con la edad cronológica de los sujetos. Terman (1943) concluye que el aumento del error a medida que aumenta el CI es una característica inherente a la técnica de la escala para obtener el CI y no se debe a otras causas, como defectos en su estructura, en la tipificación de los niveles superiores o a la inestabilidad de los sujetos. La fluctuación de un sujeto en su rendimiento está más de acuerdo con su grupo de EM que con su grupo de EC.

En el manual de la Segunda Edición, se informa del error estadístico con la expresión *Error Probable* del CI (EPci). No se puede hablar de *el* EPci, sino más bien de la distribución de los EPci, creciente a medida que aumentan los valores de CI. La valoración del EPci, contrastando los resultados de la forma L con los de la forma M (método de formas paralelas), fue en términos globales del 3 % del CI obtenido, para una fiabilidad de 0,924 en niveles medios de CI. (Terman y Merrill, 1943). Combinando dos métodos diferentes, McNemar (1942) confirma el patrón de heterocedasticidad de los errores. En el manual de la Tercera Edición (ver-

sión 1960) no se ofrecen datos de los errores o fiabilidades de las muestras utilizadas. En la presentación de las normas de 1972 no se tratan estos aspectos.

Aunque la heterocedasticidad está siempre presente, su magnitud depende de la muestra utilizada y del método empleado para su cálculo. La estimación del 3-4% puede ser significativamente diferente para otras poblaciones y otras épocas. Debe ser conocida para evaluar casos individuales y para estudios e investigaciones en las que se utilice la escala. Además de no cumplirse el supuesto de homogeneidad en la distribución de los errores (homoscedasticidad) exigido para que sean válidos los análisis ANOVA y ANCOVA, puede llevarnos a conclusiones erróneas incluso a partir de una simple correlación lineal, especialmente si sólo se trabaja con CI elevados, con un rango estrecho de puntuaciones y/o combinando diferentes EC.

### Efectos del cambio en la naturaleza del CI

Las escalas de edad están diseñadas originalmente para el uso de un CI de ejecución, no precisando en realidad de normas si los criterios de paso están ajustados adecuadamente. La Segunda Edición de la escala fue la última en utilizar un CI de este tipo, obtenido por la fórmula clásica. La única corrección es la de reducir progresivamente a partir de los 13 años la edad cronológica real antes de calcular el CI. Las tablas de la Segunda Edición simplemente ahorran calcular el cociente, incluyendo además la corrección indicada. En la Tercera Edición se abandona el procedimiento propio de las escalas de edad de ajustar los criterios de paso de las pruebas hasta la consecución de un CI de ejecución. Sus normas (las de 1960 y las de 1972) ofrecen unos CI de desviación para unos valores de EM obtenidos a través de una escala de edad tipo Binet. Las distribuciones de los valores de CI obtenidas por cada muestra de edad se normalizan, con media de 100 y DT de 16. Ahora el CI no se calcula; sólo se puede obtener consultando las normas correspondientes.

La relación ideal entre EM y CI de la Segunda

Edición se representa gráficamente con una recta para cada edad cronológica, a partir de los dos años.

torsiones de las normas de 1972 se exponen en la Tabla 2.

Tabla 2. Distorsiones de las normas de 1972 de la escala Terman

valor de m	CI < 100	Efectos sobre el CI	CI > 100
m < 0 (EC menor de 8 años)	Infravaloración		Sobrevaloración
m > 0 (EC mayor de 8 años)	Sobrevaloración		Infravaloración

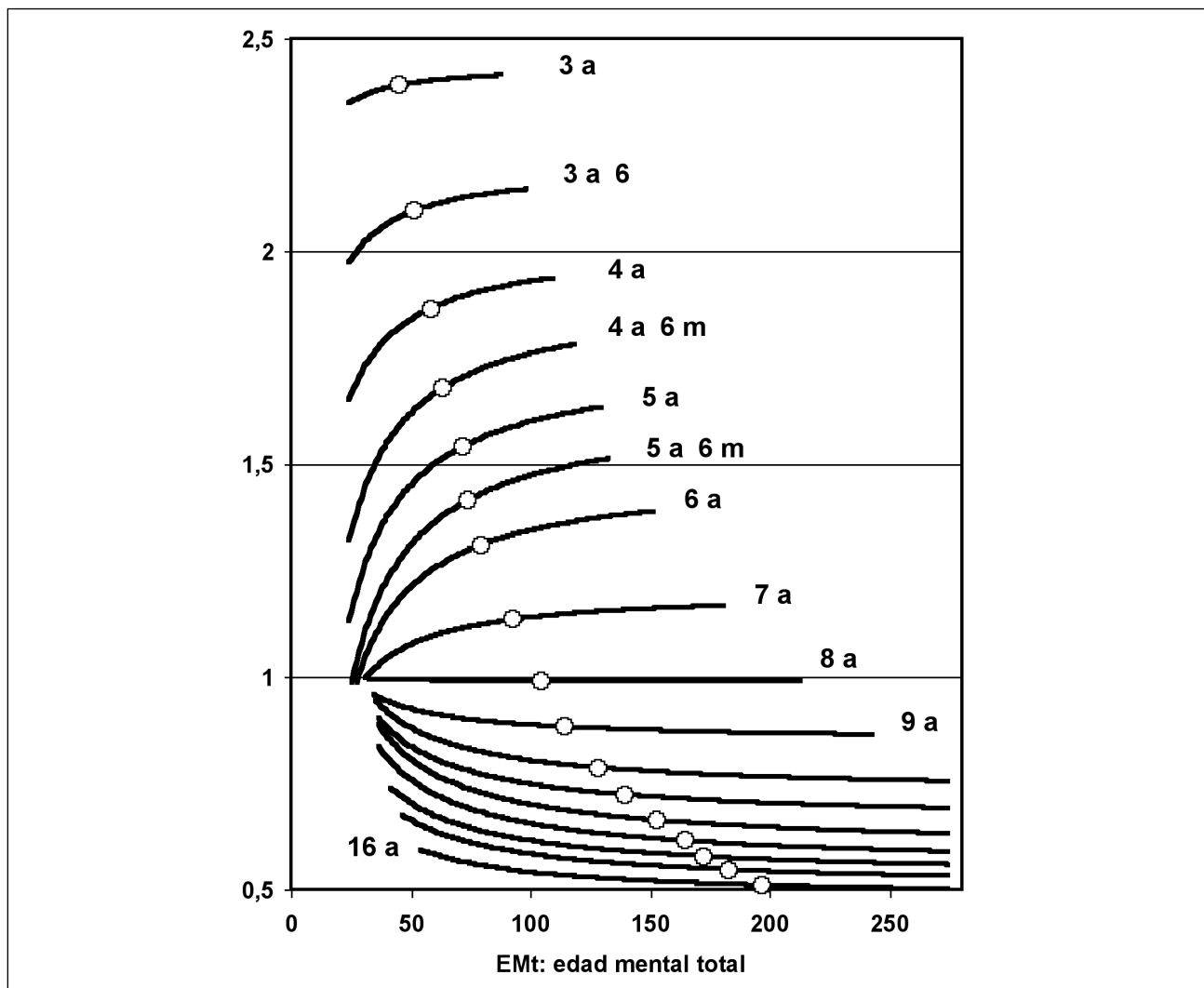
Estas rectas, que se suceden en forma de abanico, con pendientes progresivamente menores, tienen un origen común ( $EM = 0$ ,  $CI = 0$ ). En las normas de la Tercera Edición, las pendientes correspondientes a las edades cronológicas son también progresivamente menores, pero la forma de abanico característica se desdibuja debido a que las alturas de las rectas varían. Las funciones que para cada edad transforman las EM en CI también pueden expresarse según los desplazamientos de edad mental:  $CI = b \times (EM + m)$ , siendo  $m = a / b$  (ver Tabla 1). Estos desplazamientos se realizan sobre las EM, es decir, sobre los rendimientos en los que descansa directamente la normalidad de los CI. Dos muestras representativas de la misma población con dos medias de EM claramente diferentes deben tener dispersiones proporcionalmente diferentes. Los desplazamientos extremos se dan en las edades de 5 años - 6 meses y de 12 años. Si el grupo de tipificación de 5 años - 6 meses hubiera obtenido un rendimiento casi 10 meses menor, la dispersión de las puntuaciones hubiera sido proporcionalmente menor (mayor curtosis). Si el grupo de 12 años hubiera obtenido un rendimiento 17 meses mayor, su dispersión también hubiera sido mayor (menor curtosis). Para  $m < 0$  las distribuciones de EM resultantes tienen mayor dispersión de la debida (menor curtosis), por tanto se sobrevaloran progresivamente los rendimientos por encima de la media y se infravaloran progresivamente los rendimientos por debajo de la media. Para  $m > 0$  las distribuciones resultantes tienen menor dispersión de la debida (mayor curtosis). En estos casos, los efectos se invierten. Los efectos de las dis-

El valor obtenido por mes de EM depende del total alcanzado en la escala (excepto para edades cercanas a los ocho años). Así, un sujeto de 6 años con un rendimiento de 55 meses (EM), obtiene por cada uno de ellos 1,24 puntos de CI (CI total de 68), mientras que a otro de la misma edad, con rendimiento de 98 meses (EM) le corresponde una valoración de 1,34 puntos de CI por cada mes de EM (CI total de 132). La Gráfica 1 presenta la distorsión de los valores de conversión de la EM en CI para cada EC, entre los 3 y los 16 años, siendo  $V_{ci} = CI / E_{mt}$  (edad mental total). Las marcas en cada línea de EC indican el CI de 100. En una escala de edad ajustada cada EC quedaría representada por una recta horizontal constante (unidades de EM de valor constante para cada edad cronológica).

Por otra parte, el valor de los meses de rendimiento (EM) deja de ser progresivamente menor con la edad cronológica, llegando incluso a invertirse para edades cercanas. Así, los sujetos de 5 años y 6 meses de edad que obtiene un rendimiento menor de 49 meses (CI menor de 64) reciben menos valor de CI por cada mes de EM que los sujetos de 6 años de edad cronológica que puntúan por encima de 74 meses (CI mayor de 96). Este efecto es incluso mayor, entre los 3 y los 6 años, si consideramos edades cronológicas más próximas.

Hasta la Segunda Edición, la relación entre EM y CI es lineal y además estrictamente proporcional. En la Tercera Edición, la relación entre EM y CI sigue siendo lineal, pero pierde la proporcionalidad. La consecuencia es que las unidades de edad mental dejan de ser iguales y pasan a tener un valor relativo

Gráfica 1. Valores de conversión de la EM en CI, para las EC de tres a dieciséis años



Nota: las marcas indican el CI de 100

al rendimiento total. La normalidad de las distribuciones del CI para cada edad ya no descansa sobre los supuestos enunciados por Terman (1943) de igual valor de las EM para cada EC y de disminución progresiva del valor de las unidades de EM a medida que aumenta la EC.

En la medida en que el grado de dificultad global de la escala haya ido disminuyendo, el efecto de las sobrevaloraciones habrá ido aumentando y el de las infravaloraciones reduciéndose. Esto es lo que parece haber sucedido cuando se comparan los resultados de la escala con los de pruebas de inteligencia actuales (Becker, K.A. 2003. p. 10).

### Dificultades de integración de la escala en el contexto del diagnóstico psicopedagógico

Una de las características de la escala Terman, como escala de edad, es la de ofrecer un valor general de CI. Sus creadores nunca pretendieron otra cosa. Se negaron a considerar o tomar partido por teorías de organización de las capacidades intelectuales en hipotéticas facultades específicas, para desarrollar sus pruebas. En su opinión, las teorías indicadas generaban más confusión y complicaban la medición del CI (Terman y Merrill, 1975). Para describir un perfil de capacidades inte-

lectuales hay que recurrir a pruebas factoriales y ajustarse al esquema teórico con el que se han diseñado y ofrecen los resultados. Observando al sujeto mientras realiza las pruebas podemos extraer una gran variedad de informaciones, incluso relevantes, para su mejor conocimiento. Pero esto no implica que del resultado de la escala o de su proceso de corrección se puedan sacar otros datos cuantitativos válidos diferentes del CI general. Ni siquiera de la dispersión de los aciertos que cada sujeto obtiene, en ocasiones a lo largo de muchos niveles de edad mental (*scatter*), se han podido extraer informaciones relevantes (McNemar, 1942). El comunicar la edad mental alcanzada por el sujeto, además de su CI, resulta redundante. De la escala sólo se obtiene una edad mental: la de la puntuación total que conduce al correspondiente CI. Edad cronológica, edad mental y CI forman un *triángulo rígido*: con cualquiera dos de los datos se obtiene el tercero.

A pesar de las advertencias de sus creadores, en ocasiones algunas de sus pruebas se utilizan como indicadores de capacidades concretas de tipo verbal o numérica. La prueba de *Vocabulario* es, con diferencia, la más empleada con esta finalidad, probablemente porque su valoración, que se extiende a lo largo de nueve niveles de edad mental (aunque solo en años pares), le da apariencia de subescala. Se informa del último nivel de edad mental en el que se supera el criterio de paso de la prueba, considerando que éste es la EM específica de *Vocabulario*. Cuando se afirma, por ejemplo, que un sujeto de 9 años tiene una EM de 12 años para la capacidad de *Vocabulario*, lo que se está indicando en realidad es que ha realizado la prueba de *Vocabulario* con un rendimiento por el que ha superado el criterio de paso correspondiente a los 12 años. Esta información es tan cierta como compleja de interpretar y valorar. Las escalas tipo Binet están construidas de forma que, en general, salvo para los niveles de adultos, las pruebas situadas en un nivel de edad mental son superadas por más del 50% de los sujetos de esa misma edad cronológica. En muchos casos, por un porcentaje ampliamente mayor

(Terman y Merrill, 1975, Apéndice B). Si las pruebas se superaran sólo por el 50% de los sujetos de la misma edad, sencillamente la escala no se podría aplicar (McNemar, 1942). La prueba de *Vocabulario* no es una excepción. Del apéndice B citado, podemos obtener los porcentajes de paso para cada edad, calculados para la versión de 1960 (últimos conocidos). Estos porcentajes de paso son, para las edades de 6, 8, 10, 12 y 14 años sucesivamente de 66, 62, 54, 65 y 77. Así, los sujetos de 12 años de edad superan en un 65% la prueba a su nivel base. ¿Qué interpretación práctica le damos a esta información? Desde luego no podemos decir que el sujeto de nueve años tiene una EM de 12 años en la prueba de *Vocabulario*. En las escalas factoriales se puede hacer referencia a un nivel de edad equivalente (EE) para cada una de las pruebas. No tiene sentido para el total de la escala, ya que para ello está el CI total. Se obtiene esta información por un procedimiento que podría llamarse *lateral*, al no ser necesario para el cálculo de los perfiles ni de los datos de CI. La edad equivalente del rendimiento de un sujeto en una prueba se obtiene comparando su rendimiento con los valores medios (medianas) de cada grupo de edad. La EE tiene un significado propio y complementario del CI.

Otra dificultad de importancia para la evaluación de capacidades es la de tratar de comparar pruebas, incluso cualitativamente, de diferentes escalas que reciben el mismo nombre, pero no implican las mismas capacidades ni se valoran con los mismos criterios. Así en *Vocabulario* de la escala Terman se valora exclusivamente el conocimiento suficiente de las palabras. El examinador ayuda al sujeto a explicarse, ya que la finalidad es la de determinar si las conoce o no. En *Vocabulario* de las escalas Weschler, además del conocimiento se valora la precisión y claridad de la definición o explicación que se hace de ellas, en una escala de 0 – 1 – 2. Evidentemente, en la prueba *Vocabulario* de las escalas Weschler el sujeto debe movilizar más recursos que en la homónima de la escala Terman. El profesional, contrastando los manuales de una y otra prueba, podrá también hacerse una idea de las dife-

rencias cualitativas y cuantitativas de las valoraciones en ambas pruebas.

### **Dificultades de comprensión del concepto de “edad mental” para los destinatarios**

La edad mental es la unidad de medida de rendimiento empleada por la escala Terman. El sujeto puede rendir con niveles diferentes (de hecho es lo que habitualmente sucede) en otras pruebas intelectuales, o que pretenden medir otros aspectos del desarrollo de naturaleza diferente (Terman y Merrill, 1943). El sujeto obtiene un rendimiento medido en meses de EM para poder comparar su adelanto o retraso en relación a otros de su misma edad cronológica, no *tiene* o *posee* una Edad Mental. El rendimiento en una prueba se suele elevar a cualidad general del sujeto por los destinatarios de la información. Considero este equívoco una de las mayores trampas del lenguaje producidas en el campo de la Psicología. Resulta insuperable para la mayoría de los padres que han recibido una valoración alta de la capacidad intelectual de sus hijos con la escala Terman. Incluso frente a un auditorio de buen nivel de preparación, expresiones del tipo “tal sujeto de 5 años y 11 meses de edad cronológica posee una edad mental de 8 años y 2 meses” (CI de 134 según normas norteamericanas de 1972)

producen, por su equívoca interpretación un impacto inadecuado y difícil de corregir. Desafortunadamente, la superación de esta dificultad no pertenece al campo psicométrico sino al de la pedagogía de los destinatarios. Solucionarla queda en manos de los que siguen empleando profesionalmente la escala.

### **Referencias**

- Becker, K.A. (2003). *History of the Stanford-Binet intelligence scales: Content and psychometrics. (Stanford-Binet Intelligence Scales. Fifth Edition Assessment Service Bulletin No. 1)*. Itasca, IL: Riverside Publishing.
- McNemar, Q. (1942). *The revision of the Stanford-Binet Scale. An Analysis of Standardization Data*. Houghton Mifflin Company. Boston.
- Terman, L. M. y Merrill, M. A. (1943). *Medida de la Inteligencia. Método para el empleo de las pruebas del Stanford-Binet nuevamente revisadas*. Traducción y adaptación de Germain, J. Madrid: Espasa-Calpe S.A.
- Terman, L. M. y Merrill, M. A. (1975). *Medida de la Inteligencia. Método para el empleo de las pruebas del Stanford-Binet. Tercera revisión Formas L y M reunidas*. Traducción y adaptación de Germain, J. (1975). Madrid: Espasa-Calpe S.A.

Artículo recibido: 04/08/2009

Revisión recibida: 02/10/2009

Artículo aceptado: 13/10/2009