

8. APLICACION PSICOMETRICA: NOTAS SOBRE EL DESARROLLO HISTORICO DE LOS TESTS MENTALES.

Trabajos sobre la historia de los tests. El desarrollo histórico de los tests mentales puede encontrarse sumariamente tratado en algunas notas más o menos extensas que suelen acompañar a los manuales de historia de la psicología (p.e. Carpintero, 1.978), a los apuntes docentes de la materia (p.e. Yela. 1.977; 1.984), a los manuales generales acerca de contenidos psicométricos (p.e. Anastasi, 1.968), o a los manuales acerca del psicodiagnóstico (p.e. Silva, 1.982). Con alguna excepción como la obra de DuBois (1.970), y algunas perspectivas como la de Van der Linden (1.986) y la de Thompson y Sharp (1.988), quizás no se ha prestado una atención suficiente a la elaboración de una historia de los tests mentales y su uso en psicología. Esta ausencia es una queja común que frecuentemente se escucha también en disciplinas vecinas como el psicodiagnóstico. Sin embargo, si de alguna parte de la psicometría se hace mención en todas estas aproximaciones, y si alguna imagen ha quedado para el resto de la psicología y para el público de lo psicométrico, es desde luego la referida a los tests mentales, entendidos mediante tres o cuatro hechos aislados, emblemáticos, que se han convertido en el estereotipo de la historia de la psicometría. Esos hechos incluyen inevitablemente las medidas sensomotoras de Galton, el test de Binet y su revisión de Stanford y la aparición del test alfa durante la Primera Gran Guerra.

Puede encontrarse una descripción detallada del momento del origen de los tests mentales en el trabajo de Goodenough (1.949) y también en el libro de Peterson "Early conceptions and tests of intelligence" (1.925). También se presenta una visión general en los trabajos de Murphy (1.949) y de Boring (1.950). Recientemente Paul McReynolds y Klaus Ludwig (1.987) en un breve trabajo se han ocupado de apuntar algunas notas acerca del origen de las escalas de medida con varias alternativas, considerando algunos antecedentes muy remotos y otros más próximos ubicados en los últimos siglos. DuBois (1.970), presenta explícitamente una historia desde el punto de vista de los tests mentales. También Thompson y Sharp (1.988) han presentado una historia de los tests mentales, muy centrada sobre una limitada selección de hechos. Las periódicas

revisiones del *Annual Review of Psychology* contribuyen a obtener un seguimiento adecuado del desarrollo histórico de la disciplina.

Los apuntes históricos que siguen constituyen tan solo una selección de hechos e interpretaciones. Una consideración más pormenorizada puede obtenerse en la cronología que adjuntamos y en parte de la bibliografía citada. Por supuesto no volveremos aquí sobre la historia de la teoría psicométrica que hemos tratado en los puntos anteriores, limitándonos a la historia de los tests mentales, a algunos aspectos institucionales destacados y a algunas consideraciones críticas. Tampoco vamos a profundizar en el material psicométrico que se menciona, limitándonos a ofrecer un panorama general.

Orígenes de los tests: Galton y MacKeen Cattell. El comienzo de los tests mentales, dejando a parte algunos antecedentes más o menos próximos, puede situarse, paralelamente a la fundación de la Psicología, en torno a la década de los ochenta del siglo pasado. En este tiempo una serie de preocupaciones sociales y educativas, y un conjunto de influencias que incluía el empirismo inglés y la obra de Darwin, aparecida a mediados del siglo, impulsaron la idea de que las diferencias individuales de los sujetos debían ser consideradas para tomar decisiones acerca de los mismos. En esta época aparecen las primeras mediciones educativas, las primeras mediciones con propósitos de selección de personal, y las primeras mediciones de carácter sensorial y motor.

Si Wundt estaba orientado hacia una concepción estructuralista de la psicología orientada hacia el hallazgo de leyes generales donde las diferencias entre sujetos eran un factor impertinente, Sir Francis Galton por el contrario, desde una perspectiva naturalista y un espíritu inquieto, buscaba precisamente esas diferencias individuales como expresión de la variabilidad humana en el curso de la evolución. A tono con el espíritu de la época, Galton sostenía la tesis hereditarista y su preocupación estuvo guiada en buena parte por el deseo de mostrar como la inteligencia tenía una razón y una determinación genética. En torno a 1.880 no había un modo establecido para medir la inteligencia de los sujetos, lo que no significa que no hubiera un diagnóstico de incapacidad mental basado en las apreciaciones personales sobre todo de médicos. Sin embargo, la tradición empirista de Locke había puesto la base de la ciencia y del conocimiento en la captación sensorial empírica, y, bajo esta apreciación, la hipótesis de que los sujetos más capaces

intelectualmente serían también los más capaces sensorialmente estaba servida. La ventaja de poner la inteligencia en lo sensorial y en la habilidad motora que dependía también de aquella, residía en la facilidad de recoger información cuantitativa y objetiva sobre la misma, fuera de vaguedades especulativas. Galton se ocupó en su laboratorio antropométrico, como es sabido, en recoger la primera masa de datos importante sobre capacidades sensoriales y motoras simples. La importancia de Galton no reside desde luego en los "tests" sensoriales y motores que usó, sino en crear la necesidad de una medición objetiva, de desarrollar métodos para el tratamiento de esa información cuantitativa, y extender este punto de vista. Al parecer Galton tuvo un lema "Siempre que puedas, cuenta" que sin duda trasmitió a la psicología que le sucedería.

Galton tuvo un discípulo principal y un mensajero y divulgador insustituible. El discípulo fue Karl Pearson, que encontró los modos de tratar con aquellos datos cuantitativos poniendo los cimientos estadísticos para utilizar la información de los tests. El gran mensajero de Galton fue James MacKeen Cattell, el creador de la expresión "test mental". MacKeen Cattell había hecho su tesis doctoral con Wundt, bajo el ceño fruncido del maestro que no veía con demasiados buenos ojos la atención del discípulo a las diferencias individuales. En el camino de vuelta a Estados Unidos, Cattell se detuvo una temporada junto a Galton. Fue suficiente para importar un conjunto de procedimientos de medición que J.M. Cattell aderezó con sus aprendizajes en Leipzig. Durante las dos décadas siguientes Cattell y sus discípulos hicieron innumerables mediciones que nunca llegaron a correlacionar con los logros educativos de los sujetos. El procedimiento era la medición pero el objeto de la misma debía ser otro. Yela (1984) ha denominado a esta época de pruebas simples, individuales y poco o nada relacionadas con los criterios educativos, periodo de los tests sensoriales y motores.

La escala de Binet y Simon. Faltaban cinco años para acabar el siglo cuando Binet y Henri sugieren que deberían medirse directamente facultades superiores. Pasaban cinco años del comienzo de siglo cuando, impulsados por necesidades educativas prácticas, Binet y Simón daban a luz la primera escala de inteligencia de complejidad moderna. Mientras Ebbinghaus había acumulado pruebas que sugerían que la intuición de Binet iba en el camino adecuado. La escala estaba basada en la observación de que determinadas tareas comienzan a resolverse a ciertas edades, de este modo las tareas que típicamente se

resuelven a una determinada edad son representativas de ese *nivel mental*. Si un sujeto resuelve correctamente las tareas de una edad superior puede estimarse en que grado está adelantado y si sólo resuelve las de una edad anterior en que grado está retrasado.

La escala de Binet aportaba una orientación hacia las tareas complejas, abandonando el camino de las tareas simples e carácter perceptivo o motor. Binet pretendía de este modo evaluar directamente los productos de la inteligencia a un nivel superior, dejando a lado una orientación analítica más orientada a ahondar en los componentes y los procesos de la inteligencia. No obstante, Binet tomó esta orientación por las necesidades prácticas de clasificación educativa a las que había de dar respuesta; sin embargo, primero se había ocupado durante diversos periodos de explorar otras aproximaciones al problema; había probado con tests sensoriales y mediciones antropométricas e incluso se había ocupado de la quiromancia.

La escala de Binet también aportaba un procedimiento para establecer el rendimiento normal de los sujetos y unos criterios para tomar decisiones pedagógicas socialmente útiles. La administración de la prueba era de carácter individual y la interpretación de las respuestas como aciertos o errores puede considerarse objetiva. No tenía ni una teoría de la inteligencia, ni una teoría de la medición bien formadas para sostenerse. La teoría psicométrica, como ya hemos tratado anteriormente, la aportaría Spearman en un conjunto de trabajos publicados en torno a estas fechas. De ahí en adelante la teoría psicométrica crecería intensamente, con una justificación útil, y la aplicación de los tests mentales se extendería rápidamente, con un soporte formal bajo sus pies.

Revisiones y aparición de otros tests de inteligencia. Las revisiones de la escala de Binet marcan el ritmo del crecimiento de la aplicación psicométrica en estos primeros años. En 1.908 y otra vez en 1.911, año de la muerte de Binet, la escala original es revisada por sus autores. Ese mismo año Healy-Fernal presenta tests manipulativos orientados a la medición de capacidades superiores, y al año siguiente Stern sugiere el famoso concepto de "C.I." o "cociente intelectual". Al parecer, Kuhlman sugirió simultáneamente con Stern este cociente entre la edad mental de Binet y la edad cronológica como indicador de la inteligencia de los sujetos. Durante esos años estas primeras escalas son traducidas a diversos idiomas europeos, y comienzan a extenderse a ambos lados del Atlántico.

Otros tests manipulativos aparecen también durante estos años, pero lo más notorio será la revisión que Terman efectuó de la escala de Binet en 1.916 y la aparición de los tests colectivos al año siguiente. El dos de abril de 1.917 el Congreso de los Estados Unidos decidió participar en la primera guerra mundial dando lugar a urgentes necesidades de clasificación y ubicación de reclutas. A raíz de este hecho histórico las pruebas de capacidad mental debieron volverse necesariamente colectivas. Quizás no se ha resaltado lo suficiente en este hecho, por lo demás bien documentado, el papel que jugaron una serie de tests que Arthur S. Otis tenía preparados con anterioridad. En realidad los tests de Otis fueron el fundamento de los elaborados por el comité presidido por Yerkes. Además de impulsar los tests colectivos se impulsaron también los tests no verbales con el desarrollo del test Beta, la aplicación a la selección y clasificación de personal y el prestigio social de los tests. Las necesidades escolares habían llevado a la primera escala de inteligencia individual, las necesidades militares llevaron a la primeros tests colectivos verbales y no verbales. El trabajo es tan intenso que en 1.918 ya es posible la aparición de una revisión de tests educativos y psicológicos. Al año siguiente aparecerá la escala de Gesell. En 1.928 Germain y Rodrigo adaptan el test de Binet en versión de la revisión de Stanford.

El test de Binet seguiría revisándose una serie de veces. Terman y Merrill en 1.937 efectuaron una de las más importantes revisiones del mismo, llevando a cabo una de las tipificaciones más cuidadosas de la historia de los tests y obteniendo dos formas paralelas. En 1.960 se publicaba la tercera revisión del Stanford-Binet con una sola forma y nuevos baremos. En 1972 se realizó un nuevo trabajo de normalización del Stanford-Binet: El resultado más sorprendente fue que en cada edad cronológica la media de los CI estaba ahora por encima de 100, oscilando entre los 101'9 de los 10 años y los 110'7 de los 3 y los 4 años. Es decir, en los años transcurridos desde 1937, los items se habían vuelto "más fáciles" para los sujetos, o dicho de otro modo, la capacidad de las muestras había mejorado, probablemente debido a efectos ambientales, educativos y de salud. A pesar de las sucesivas revisiones de la escala de inteligencia más estudiada, sorprende que todavía recientemente se califique de insuficiente la información sobre la fiabilidad y la validez de las últimas versiones (Graham y Lilly, 1.984). Por otra parte, no debe perderse de vista que los datos para fiabilidad y validez pueden diferir -y difieren de hecho- para diferentes edades.

El año 1.939 aparece el test de Weschler-Bellevue para la medición de la inteligencia de los adultos; en 1.950 aparecerá

una extensión para niños denominada abreviadamente WISC. Los tests de Wechsler comprenden fundamentalmente el WAIS, para adultos, el WISC, para niños y adolescentes, y el WPPSI, para pre-escolares, tratándose de las escalas de inteligencia de aplicación individualizada probablemente más utilizadas. A diferencia de la escala de Stanford-Binet, las escalas de Weschler permiten obtener separadamente un CI verbal y un CI manipulativo, obtenidos a partir de un conjunto de subtests, además del CI general.

Cuestionarios de Personalidad. En 1917, al mismo tiempo que los primeros tests colectivos de inteligencia, se desarrolló la hoja de datos personales de Woodworth, que es citada a veces como el primer test de personalidad. En 1.942 aparecerá la primera versión del M.M.P.I. (Inventario Multifásico de Personalidad de Minessota) de Hathaway y McKinley, un test que todavía actualmente es objeto de una intensa investigación en el campo de la personalidad, clínico y patológico. El campo de los cuestionarios o inventarios de personalidad se ha desarrollado intensamente desde entonces, sin embargo, el MMPI, junto con el 16 PF (Cuestionario de Personalidad de 16 factores) de Raymond B. Cattell, el Inventario Psicológico de California de Gough (1975), y el Inventario de Preferencias personales de Edwards (1959), constituyen algunos de los que parecen resultar más utilizados internacionalmente. En Europa destaca el trabajo desarrollado por Eysenck y sus colaboradores cuyo fruto más conocido es el EPI (Eysenck Personality Inventory) y el EPQJ (Junior Eysenck Personality Questionnaire) muy ampliamente utilizados. Otros cuestionarios como el BELL, el CEP y el Guilford-Zimmerman también gozan de un uso frecuente.

Tests Proyectivos. Durante la primera mitad del siglo además de los tests psicométricos habían ido apareciendo también un conjunto diverso de tests que a veces se denominan bajo la común etiqueta de proyectivos. En 1.921 el "Rorschach", en 1.926 el de dibujo de la figura humana, en 1.938 el TAT. La idea esencial de estas pruebas es muy sugerente: utilizar estímulos ambiguos de modo que la variabilidad de las respuestas de los sujetos únicamente puedan atribuirse a los sujetos, a su "proyección" sobre el estímulo. Cronbach en 1.949 efectuó una revisión de los trabajos que aportaban resultados empíricos sobre estos tests que ponía en evidencia esos estudios. El trabajo de Cronbach puede tomarse como punto de referencia en el desarrollo de una literatura que vendrá a concluir

insistentemente que los procedimientos "clínicos" carecen de la fiabilidad y la validez necesarias, aun cuando, según sus defensores, aporten fuentes de información relevante. En esa línea crítica cabe considerar el trabajo de Meehl de 1.954. A pesar de estas críticas, especialmente el test de las manchas de tinta de Rorschach ha ejercido un particular poder de fascinación sobre muchos psicólogos dando lugar a una gran cantidad de investigaciones -en su mayoría críticas-, a una serie de sistemas no siempre coincidentes de puntuación e interpretación -entre los que puede destacarse el de Exner-, y a un extenso uso y abuso profesional.

Desarrollos institucionales. Paralelamente a lo que es el desarrollo del método práctico de los tests hay que considerar los avances y la consolidación institucional que la disciplina ha ido adquiriendo. En 1.922 J.M. Cattell fundó la primera empresa dedicada a la publicación masiva de tests y a partir de ese año, debido a un cambio de dirección, los tests comienzan a ser la herramienta corriente de selección en la administración americana. En 1.935, también dentro de los acontecimientos institucionales, debe mencionarse la creación de la "Psychometric Society". Al año siguiente se producirá la abolición de los tests mentales en la URSS y la primera edición de los "Mental Yearbook" de Buros.

En 1.947 tiene lugar un hecho institucional que ha tenido gran importancia en la historia de la medición psicológica. Se trata de la fundación del "Educational Testing Service". El ETS ha jugado un papel muy importante en el desarrollo histórico de la psicometría como soporte y núcleo de formación de buena parte de los principales psicómetros que aparecen en las décadas siguientes.

En 1.954 aparecen las recomendaciones técnicas para tests de la APA. Una publicación que constituye un punto de referencia común, renovado en ediciones sucesivas en 1.966, 1.974 y en 1.985. En 1.969 la APA editaría también recomendaciones acerca del examen psicológico de personas poco dotadas.

Críticas. Durante las década de los cuarenta y los cincuenta los tests habían llegado a un nivel de uso, y también a ciertos abusos, que los volvía fácil blanco de las críticas sociales. Yela (1.977) relata que "solo en los Estados Unidos se aplicaron durante 1.946, unos sesenta millones de tests a unos veinte millones de personas." Durante la segunda parte de los años

cincuenta comenzó a generarse en Estados Unidos un malestar social acerca del uso de los tests que se plasmó en diversos hechos espectaculares, como la cremación pública de tests en Texas o la prohibición de los tests en California en 1.959, o la prohibición de la administración de tests en las escuelas neoyorquinas en 1.964. En 1.965, mientras que el Congreso estadounidense prohibía los tests de personalidad en la selección de empleados federales, un comité de la APA con espíritu más constructivo llegaba a una serie de recomendaciones éticas y prácticas que aparecieron en el *American Psychologist*. Sin embargo, a veces se ha exagerado mucho la cuestión hablando de una "revuelta anti-test".

Las críticas sociales han ido acompañadas de una serie de desarrollos críticos en la literatura científica no exentos de ciertas polémicas. Desde el diagnóstico, como en el trabajo editado por Pawlik en 1.976 por ejemplo, vendrán críticas acerca del lugar y la capacidad de los tests. El desarrollo del concepto de adecuación de los tests, que puede considerarse que aparece en los trabajos de Levine y Rubin de 1.979, y el énfasis en la cuestión de los sesgos de los tests, son un exponente de un clima de preocupación por la idoneidad de los tests para determinados usos y ante determinados colectivos.

Un antecedente importante de las críticas que la teoría de los tests ha recibido desde su propio ámbito teórico puede encontrarse en un trabajo de C. H. Coombs publicado en 1.950 en el *Educational and Psychological Measurement* (Vol. 10; págs 43-56). En este trabajo Coombs defiende que la teoría de los tests psicológicos carece de una fundamentación psicológica, considerando además el procedimiento de estímulo único característico de la construcción de tests, como inadecuado para el establecimiento de la fiabilidad. La crítica de Coombs es del todo constructiva y el mismo ensaya en este trabajo una fundamentación psicológica de algunos conceptos como homogeneidad y fiabilidad.

Quizás la obra más importante acerca de la consideración de los sesgos de los tests en función del sexo, la raza, o la condición social es la de Arthur Jensen "Bias in mental testing" (Sesgos de los tests mentales) publicada en 1.980; aunque fue una pequeña monografía que publicó en 1.969 la que hizo correr mucha tinta en medios científicos y no científicos. En lo que se refiere a los tests, la tesis de Jensen, para la que aporta un buen número de pruebas empíricas, es que no puede afirmarse que los principales tests de medición de la inteligencia estén sesgados contra minorías raciales o económicas. Jensen ha dedicado tres capítulos de su obra de 1.980 a sopesar el

movimiento social ante los tests y también las feroces críticas de que ha sido objeto. Ejemplos de esa crítica no muy ajustada a los límites de la contrastación empírica es la obra de Kamin de 1.971 y, mucho más exacerbada, la de Gould "The mismeasure of Man".

Con un estilo sarcástico y metafórico, pero con un tono y una dirección muy distintas a las de las críticas anteriores, debe mencionarse la revisión de James Lumsden, recientemente fallecido, sobre teoría de tests publicada en el Annual Review of Psychology en 1.976. La atípica revisión de Lumsden es una meditación en voz alta, sincera y desencantada, acerca de una parte del desarrollo de la teoría psicométrica. Sin embargo, aunque Lumsden no hace un esfuerzo por reformular positivamente aquello que critica, la conclusión que puede extraerse de su trabajo apunta en la dirección de que determinados supuestos clásicos sobre tests deben ser revisados, más que deba abandonarse la investigación en el campo. En ese sentido las críticas de Lumsden incardinan con las de Coombs y con otras, como las de Guttman, fundadas en una necesidad de crecimiento y renovación teórica.

La cuestión puede consistir en no confundir la teoría psicométrica con los tests derivados bajo ella, ni los tests con las teorías psicológicas (p.e. heredabilidad o diferencias intelectuales raciales), ni los tests con los usos inadecuados de los mismos (p.e. en diagnóstico o en selección con fines discriminativos). El propósito de la medición psicológica, tal como nació a principios de siglo es básicamente técnico y socialmente positivo. Esos propósitos y directrices éticas son distintas de la teoría psicométrica, de la aplicación psicométrica y de los usos psicométricos, pero deben guiar obviamente la utilidad social del trabajo psicométrico.