

2024



ELEMENTOS DE PSICOMETRÍA

UN BOSQUEJO HISTÓRICO



Adolfo López Suárez
Francisco Javier Galatuart Alonso
Karen Janeth Gutiérrez Hernández
Ana María Ibarrola Martínez
Jenny Javier Hernández





Integrantes del seminario:

Adolfo López Suárez
Francisco Javier Galatuart Alonso
Karen Janeth Gutiérrez Hernández
Ana María Ibarrola Martínez
Jenny Javier Hernández
Samantha Paulina Camacho Portillo
Ian Carlo Guzmán De La Riva
Axel Isaí Urrutia Reyes

Diseño editorial:

Víctor Luna Lomelí

Trabajo producido por el Seminario Permanente de Psicometría, dirigido por el Dr. Adolfo López Suárez en el Instituto de Estudios Sobre la Universidad, de la Universidad Autónoma del Estado de México.



*“ Si uno empieza con certezas, acabará con dudas.
Pero si se empieza con dudas, se logrará llegar a certezas,
aún si faltan las palabras para expresarlas.”*

HERÓDOTO DE HALICARNASO

CONTENIDO

4	<u>Presentación</u>
5	<u>China</u>
9	<u>Grecia</u>
13	<u>Roma</u>
15	<u>Edad Media</u>
18	<u>Edad Moderna</u>
22	<u>Siglo XIX</u>
27	<u>Siglo XX</u>
35	<u>Siglo XXI</u>
39	<u>Conclusiones</u>
42	<u>Cronología</u>
44	<u>Bibliografía</u>



Presentación

Pitágoras de Samos, filósofo presocrático, postuló que todo es número. Más de dos milenios después, William Thomson, mejor conocido como Lord Kelvin, en el siglo XIX afirmaba que cuando se puede medir aquello de lo que se habla y expresarlo en números, entonces se sabe algo al respecto; cuando no se puede medir ni expresarlo en números, entonces el conocimiento es escaso e insatisfactorio; Kelvin también pensaba que la ciencia comienza cuando se comienza a medir.

Es probable que la razón principal por la que la psicología surgió como una ciencia hasta 1860 sea la dificultad, hasta entonces insalvable, para medir los fenómenos psicológicos.

Como ocurre con cualquier otra ciencia, la física incluida, el avance de cada disciplina ha seguido al desarrollo de instrumentos de medición cada vez más potentes y precisos. No se concibe la biología actual sin el microscopio, la astronomía sin el telescopio o la física sin el gravímetro; las definiciones mismas de constructos fundamentales en las ciencias más objetivas han estado condicionadas a los instrumentos de medición disponibles. El caso de la psicología no puede ser la excepción a esta regla.

Pero aquí es más complejo. A diferencia de lo que ocurre en otras ciencias, los objetos de estudio de la psicología no pueden someterse a observación directa. Desde el fenómeno más elemental, la sensación, hasta la inteligencia, la memoria o la angustia, estamos ante constructos que, para decirlo claramente, tienen que ser inferidos a partir de conductas abiertas que emite el sujeto bajo estudio. Desde el origen, es fácil comprobar esto si se analizan los experimentos de Weber y Fechner.

Si conceptuamos a la psicometría como la disciplina que se ocupa de la medición de los

fenómenos psicológicos -concepto que resulta muy general y con múltiples aristas que deben ser analizadas-, entonces puede afirmarse que la psicología científica nace junto con la psicometría. Es cierto que comúnmente el término psicometría se relaciona con los tests, pero disociar a este tipo de instrumentos de otros, como los psicofísicos, resulta realmente reduccionista y muy pronto conduce a contradicciones.

A partir de estas consideraciones, un grupo de psicólogos y estudiantes decidimos iniciar un seminario para estudiar temas selectos sobre la psicometría. Se pretende producir una serie de fascículos que apoyen la difusión de las teorías, métodos e instrumentos que se han desarrollado para medir los fenómenos psicológicos.

Este es el primer fascículo y en él se presenta una revisión histórica del desarrollo de la psicometría, desde los primeros intentos para medir aptitudes con fines de selección, realizados hace más de 2,000 años, hasta algunos de los desarrollos de frontera del siglo que corre. El propósito de este trabajo consiste en invitar a reflexionar sobre la evolución de la medición en la psicología científica. Hemos tratado de ubicar con la mayor precisión posible algunos de los avances más importantes que han marcado el desarrollo de la psicometría. Para el interesado en profundizar en este fascinante campo de estudio, cada uno de los eventos que se señalan se ubica cronológicamente y se sugiere una referencia bibliográfica, lo más cercana posible al planteamiento original.



China

Más allá de estereotipos, en Occidente es muy poco lo que la mayoría conoce sobre el desarrollo de la psicometría en las culturas orientales. Esto abre el espacio al conocimiento histórico, pero también a la leyenda, y debemos estar prevenidos de tal riesgo. Existe controversia sobre la antigüedad de los exámenes como instrumentos para evaluar aspirantes a ocupar cargos públicos, lo que devendría en un servicio civil de carrera. Sin embargo, hay acuerdo en cuanto a su lugar de origen. Anastasi (1973) y DuBois (1964) proponen que hacia 2200 a.C. en China, con la dinastía Xia, se inició la práctica de evaluar las habilidades de los aspirantes como medio para

asignar puestos oficiales y que para 1115 a.C. en la dinastía Shang ya se aplicaban exámenes. Por otra parte, Bowman (1989), argumenta que no existe evidencia de lo anterior, aunque también ubica en China, el origen de los exámenes 2,000 años más tarde, hacia el 150 a.C.

En cualquier caso, se sabe con razonable certeza que desde tiempos muy remotos



Examen en Kaifeng, dinastía Song, China (960-1279) [Pintura], El país ([URL](#)). De dominio público.



competían por los cargos imperiales aspirantes provenientes de vastas regiones, y que eventualmente abarcaron a todo el imperio. Para comprender cómo fue posible la cobertura que tuvo este sistema de evaluación de aspirantes, debe recordarse que los chinos inventaron y desarrollaron la escritura logográfica, basada en caracteres llamados sinogramas, cada uno de los cuales significa una idea o un objeto; este sistema de escritura, con la natural evolución, ha permanecido vigente hasta nuestros días. Se han inventariado cerca de 50,000 sinogramas, pero se estima que alrededor de 5,000 son suficientes para la mayoría de las personas en la comunicación cotidiana. El hecho importante es que quienes dominen al menos este número de caracteres pueden comunicarse entre sí, independientemente del idioma que hablen.

Salvado en lo esencial el obstáculo del lenguaje, el sistema de exámenes chino permitía la competencia de cualquier persona que decidiera inscribirse, sin requisitos de origen social, antecedentes ocupacionales o incluso edad. Los aspirantes, que ambicionaban un puesto oficial para progresar, se preparaban a veces por décadas para presentarse al examen y los que eran rechazados podían concursar tantas veces como quisieran. Las aplicaciones de exámenes variaban en periodos de uno a tres años.

Así se abrió la posibilidad de ascenso social y económico, evitando que los puestos en el gobierno se reservaran a la aristocracia. En consecuencia, en China no se desarrolló una clase gobernante y los cargos oficiales en la mayoría de los casos se asignaban por aptitud, lo que sin duda favorecía la eficiencia de la administración pública. El sistema de exámenes colaboró, posiblemente como uno de los factores principales, a lograr la estabilidad que la sociedad china mantuvo durante más de 3,000 años. Aunque la continuidad social fue

interrumpida temporalmente por eventos catastróficos como la invasión mongola, al restablecerse el orden imperial los exámenes se reiniciaban.

En 622 d.C. se había desarrollado un complejo sistema de exámenes que cubría a todo el imperio. El objeto de evaluación se centraba en la aptitud para ejercer cargos de gobierno. Las poblaciones enviaban concursantes a su distrito; los distritos enviaban concursantes a la capital provincial y las provincias los enviaban a Pekín. Las condiciones para estandarizar el proceso de evaluación eran rigurosamente controladas. Los exámenes finales se realizaban en la Ciudad Prohibida, bajo la supervisión del emperador en persona. Así, en las diferentes etapas concursaban cientos de miles de aspirantes. La proporción de ganadores respecto al número de concursantes era tan minúscula, que no parece probable que se presentaran problemas de validez teórica. Como sucede hoy con las grandes competencias deportivas, los ganadores en las diferentes etapas, aunque no recibieran puestos o premios en efectivo, ganaban tal celebridad y respeto en su poblado, región o provincia, que el reconocimiento social compensaba con creces cualquier esfuerzo realizado. El Sistema Imperial de Exámenes para el ingreso al servicio civil en China era llamado Keju (科举)

En la actualidad, desde 1952 se aplica a gran escala el Examen Nacional de Acceso a la Universidad en China, coloquialmente llamado Gaokao (高考), que en mandarín significa "Examen Superior".

De este modo, a lo largo de milenios quienes diseñaban los reactivos que conformaban los exámenes diseñaron también el sistema educativo, económico y político; en suma, la sociedad china. No se sabe que haya llegado a



desarrollarse una tecnología para el diseño y análisis de reactivos, así como una estructura de contenidos bien definida para los exámenes. Sin embargo, se sabe que las competencias que se evaluaban eran artes y música, tiro con arco, equitación, escritura y aritmética; con el tiempo perduraron solo las dos últimas. La filosofía de Confucio (551–479 a.C.) tuvo una influencia determinante en la estructura de contenidos de los exámenes. En general, se evaluaba habilidad verbal (juegos de palabras, oratoria y argumentación), razonamiento y solución de problemas.

Se puede observar que en la historia de China la educación se orientó en gran parte a la

preparación para sustentar los exámenes. Puesto que se carecía de un sistema educativo como hoy se le concibe, cada uno se esforzaba por encontrar cómo prepararse, ya sea mediante el autoaprendizaje o bien buscando sitios donde estudiar. Esto puede explicar por qué no se desarrollaron universidades ni centros escolares en la mayor parte del imperio.

En cuanto a la validez empírica de los exámenes, se controlaba mediante el registro del desempeño de cada individuo en el puesto. Por supuesto, todo esto suponía una burocracia enorme, compleja y eficaz.

Cuando Occidente se acercó a Oriente y se interesó en estudiar su cultura, descubrió con



Examination Hall With 7500 Cells, Canton, China [Fotografía], 1873, Flickr ([URL](#)). CC BY 2.0

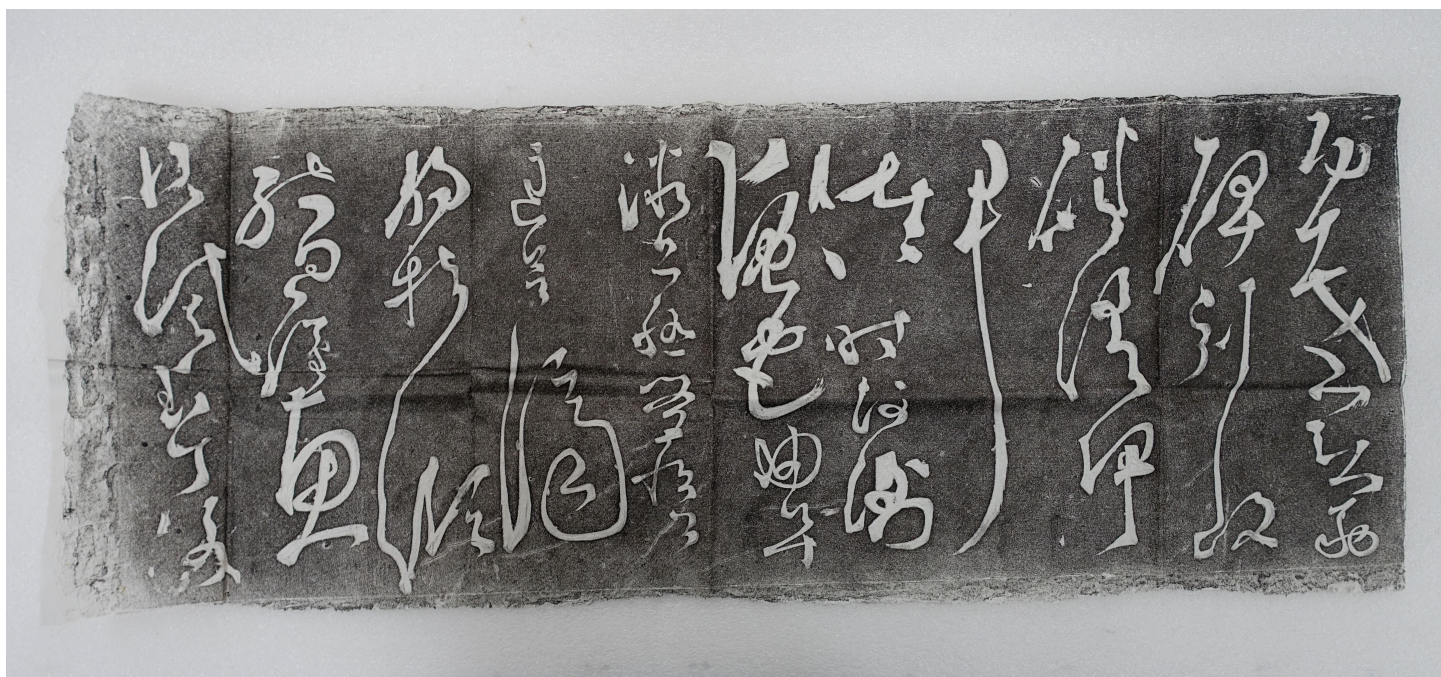


asombro su concepción del servicio civil de carrera, que aquí simplemente se desconocía. Existen registros de que en el siglo XIX los Congresos del Reino Unido y de Estados Unidos estudiaron el sistema de exámenes chino, cuando estas naciones comenzaron a diseñar sus propios servicios civiles de carrera como medio para impulsar la meritocracia y la estabilidad social (Jenckes, 1868).

Cualquiera que haya sido el año en que se originó el sistema de exámenes, éste se extendió -con sus inevitables altibajos- al menos durante 2,000 años. Es un hecho histórico que, como pocos países en el mundo, China logró durante largos periodos vivir en paz y seguridad, lo que le permitió desarrollar avances en ciencia y tecnología y una cultura única. Sin embargo,

en el siglo XIX con las invasiones, la Guerra del Opio y los boicots occidentales, descubrió que su atraso en la ciencia y sobre todo en la tecnología militar la habían dejado indefensa ante las agresiones del exterior. Obligada por ello, en 1905 abolió el sistema de exámenes y el servicio civil de carrera (Bowman, 1989). El resto de la historia es conocida y cada día lo vemos más claramente.

Si se consideran los antecedentes del sistema de exámenes chino, del que aquí apenas se trazaron algunos rasgos generales, resulta más fácil comprender el impresionante actual Gaokao, examen de acceso a la universidad, fuertemente estandarizado, que se aplica anualmente a nivel nacional durante dos días.



Fragmentary 1000-character essay written in cursive script in the Tang dynasty [tinta sobre papel]. de The Metropolitan Museum of Art, 20th century, Jstor ([URL](#)). Dominio público.



A cercándonos a Occidente, según la Biblia, cuando Israel se vio sometido y agobiado por los madianitas contaba con 32,000 hombres, pero ante un ejército enemigo muy superior estaban temerosos y no querían enfrentarlo. Inspirado por Dios, Gedeón dijo a sus hombres que podían irse quienes tuvieran miedo; solo 10,000 permanecieron. A ellos los llevó a beber para observarlos, aquellos que lamieron el agua como el perro y los que doblaron sus rodillas para beber fueron separados; solo trescientos bebieron llevando con las manos el agua a su boca. Huelga decir que con esos trescientos -y naturalmente con la ayuda de Jehová- exterminaron a sus enemigos

Grecia

(Jueces 7, 5-7). Como se ve, el método de selección involucraba la observación de la conducta, por más bizarra que ésta fuera.

En la cultura griega, interesada sobre todo en lograr el desarrollo armónico de la mente y del cuerpo, las pruebas tuvieron una gran influencia social. Los griegos se esforzaron tanto en evaluar lo físico como lo mental en el individuo, midiendo las diferentes facetas que debía poseer



Greece [Fotografía], por Moraites, P., (1870), Jstor ([URL](#)). CC BY 4.0.



el ciudadano ideal. Es precisamente de la validez de la evaluación, es decir, de la definición de lo que debía ser el ciudadano ideal, de lo que se ocupa Platón en toda su obra, pero es especialmente claro en La república y en Las leyes.

Los griegos se interesaron en la teoría política como medio para conformar la sociedad ideal que pensaron los filósofos. Los individuos debían ser cuidadosamente seleccionados y formados para ocupar los puestos en el gobierno, lo que hizo que se interesaran por un servicio civil parecido al de China, pero con las inevitables diferencias entre una sociedad que se esforzaba por ser democrática y en consecuencia era inestable y cambiante, frente a otra fuertemente autoritaria y estable. Platón reconoce las diferencias individuales y sostiene que en función de ellas deben ser seleccionados los más calificados para dirigir al Estado.

Cuando Platón se cuestionó sobre la confiabilidad de la evaluación, hizo ver que los niños siempre preferirán los títeres, los jóvenes la comedia, los adultos la tragedia y los más viejos la poesía de Homero. En consecuencia, para evaluar debían seleccionarse a los jueces más adecuados conforme al sujeto y no solo al objeto de evaluación y formar asambleas que juzgaran colegiadamente. Con esto se conformaba el método de confiabilidad interjueces.

Mediante el proceso llamado efebía, a los jóvenes de 15 a 20 años se los sometía a instrucción militar, en ocasiones en batallas reales, para adquirir la categoría de ciudadanos libres. Los aspirantes a efebo debían aplicar en un examen de estado para determinar su aptitud. En Atenas se evaluaba principalmente capacidad física y moral en pruebas individuales o colectivas. En Esparta las pruebas eran principalmente físicas, por ejemplo, los jóvenes eran flagelados para humedecer con su sangre

el altar de Diana y su motivación era tal, que preferían morir antes de decepcionar a su familia.

A diferencia de Esparta, Atenas tenía instituciones de educación superior como la Academia y el Liceo. Aunque en ellas se cultivaba el alma (filosofía), implícita o explícitamente se daba prioridad al cuerpo (gimnasia). En general, en las instituciones de educación superior se evaluaba a los aspirantes a ingresar y se rechazaba a quienes no demostraban aptitud. En la Academia, Platón examinaba a los aspirantes en matemáticas (“Que no cruce esta puerta quien no conozca la geometría”, se dice que estaba grabado en la puerta de entrada); en el Museo de Éfeso se aplicaban exámenes de cirugía, instrumentos, tesis y problemas a quienes querían estudiar medicina.

Aquí deben distinguirse dos conceptos: la aptitud es la capacidad potencial que tiene el individuo conforme a sus disposiciones personales para aprovechar el proceso educativo; en el otro extremo, la competencia es el nivel al que el individuo puede desempeñar las tareas después de un proceso de aprendizaje.

En La república, al describir al Estado ideal, Platón da gran importancia a la aptitud como criterio para seleccionar a los individuos que ocuparán los puestos de mando. Sin embargo, tanto en aptitud como en competencia deben considerarse al cuerpo y al alma, conforme las características del puesto a ocupar. De este modo, las columnas que sostenían el sistema de evaluación griego eran lo físico y lo mental, ambas estrechamente relacionadas.

Los griegos desarrollaron sistemas de evaluación altamente estructurados mediante pruebas que medían las competencias que se consideraban necesarias para desempeñar los puestos, ya fueran militares o políticos. En las pruebas griegas el problema de la validez se dejaba a los

sabios de la ciudad y el problema de la confiabilidad era responsabilidad de una asamblea de jueces, generalmente compuesta por ancianos, nombrados por la ciudad (Doyle, 1974). Como ya se dijo, los griegos daban importancia a lo físico y a lo mental.

Las pruebas de desempeño físico incluían carrera, salto, lanzamiento de disco y de jabalina, lucha y otras pruebas de destreza y resistencia. Se buscaba que los criterios de medida fueran objetivos, tales como tiempo,



*Young Spartans [Pintura], por Degas, E., 1860,
Wikipedia ([URL](#)) CC BY 4.0.*

distancia, peso, etc. Por otra parte, se trataba de estandarizar todas las pruebas para evitar injusticias en la evaluación y se controlaban rigurosamente las condiciones, como la precisión en la línea de salida y la medición de distancias; igual cuidado se ponía en cuanto a la ejecución, por ejemplo, en el salto solo se aceptaba cuando los dos pies tocaban el piso al mismo tiempo; cualquier desviación provocaba la descalificación. La preparación de los ejercicios físicos era controlada por los pedotribas (entrenadores) con el fin de estimular el máximo desempeño y de conservar la salud del cuerpo.

Una prueba clásica de psicomotricidad era la orchesis, una compleja danza que combinaba agilidad, fuerza, armonía y sentido del ritmo. La danza se ejecutaba en forma individual o colectiva en el teatro e incluso en las olimpiadas y permitía a los ejecutantes no solo expresar los sentimientos más profundos, sino también demostrar sus capacidades psicomotoras.

Platón rechaza las competencias deportivas, como las olimpiadas, porque considera que son ociosas y solo alimentan la vanagloria; en cambio, la formación física debe preparar a la juventud para la guerra, por lo que lo importante es la rapidez, agilidad y fuerza en situaciones de combate. Solo los guerreros con mayor experiencia deben juzgar a los niños. Por otra parte, en Esparta se evaluaba la aptitud del individuo desde el nacimiento; los padres entregaban a su hijo a expertos de la ciudad, quienes revisaban cuidadosamente a cada niño y los que presentaban deformidades o una constitución demasiado débil eran sacrificados lanzándolos a un abismo.

En cuanto a lo mental, se buscaba desarrollar el pensamiento analítico y racional. Las pruebas de desempeño mental evaluaban habilidades de lectura, escritura, habilidad verbal y argumentación. Para las pruebas de literatura se seleccionaba rigurosamente a los autores y se utilizaban textos no resumidos para lograr su permanencia mediante transmisión de una generación a otra. En los exámenes de lectura se hacía leer a los examinados en copias muy usadas y con la escritura sin espacios ni puntuación alguna, con el fin de evaluar la capacidad para descifrarlos y pronunciarlos correctamente apoyándose solo en su métrica. Se evaluaba la capacidad para relacionar áreas diferentes en disputas dialécticas, así como para solucionar problemas de diferentes clases. Finalmente, se consideraba importante el aspecto utilitario del aprendizaje,

pero también se apreciaba el conocimiento por sí mismo.

También se evaluaba la personalidad. Platón sostuvo que la cualificación mental del individuo debía conjugar conocimiento, buena voluntad y franqueza. Para probar el temperamento y carácter se debía observar al sujeto desde la niñez; probar su capacidad para no caer en el engaño; su coraje en ejercicios militares reales en los que hay peligro; cuando sea seguro, convenía hacerlos probar la sangre; probar su resistencia al miedo sometiéndolos al ruido y la confusión en el combate; probar su disposición a la ira, la lujuria y el orgullo mediante el vino.

Finalmente, los griegos pensaban que debe darse al sujeto poder sobre los otros oficiales en situaciones de combate, para probar la medida en que son capaces de mantener su capacidad de mando o bien para ver si flaquean. A la edad de quince años, se decía, deben otorgarse los puestos de mando del estado a quienes han superado de la mejor manera todas estas pruebas.

Es difícil saber si hubo influencias mutuas entre China y Grecia, pero es claro que los objetivos que ambas culturas buscaban, así como los procedimientos que seguían al decidir a quienes debían otorgarse los puestos de gobierno, eran similares.



Léonidas aux Thermopyles [Pintura], por David, J. L., 1814, Wikipedia ([URL](#)). Dominio público.



Roma

Según la leyenda, Roma se fundó en 753 a.C. y funcionó como una monarquía hasta 509 a.C. En este periodo se consolidó una cultura que dependía fundamentalmente de la fuerza para sobrevivir, por lo que dio prioridad al desarrollo militar. Esta característica, que se conservó hasta finales de la Edad Media, si incluimos al Imperio Romano de Oriente, dio origen al que sería el ejército que más tiempo perduró en toda la historia. En principio, todos los ciudadanos en edad registrados en el census debían cumplir el servicio militar; con excepción de la aristocracia, los demás eran seleccionados y ascendían por competencia, tomando como criterio el mérito que cada uno demostraba en la guerra.

La sociedad romana se dividía en tres clases sociales básicas. Primero estaban los patricios, los más ricos y poderosos, a quienes se

reservaban los principales cargos públicos. Después estaban los plebeyos, que eran ciudadanos libres que formaban las clases media y baja, con menores privilegios que los patricios; el estrato más bajo de los plebeyos se formaba por los proletarii, que eran tan pobres que estaban libres del pago de impuestos y su función social principal era tener hijos. Finalmente estaban los esclavos, que carecían de derechos y pertenecían a ciudadanos libres.

Durante la República (509–27 a.C.) cobró mayor importancia la ley. Formar parte del ejército representaba no solo honores, sino beneficios concretos para quienes carecían de riqueza y poder. Los aspirantes se presentaban voluntariamente, incluso llegaban a competir por los puestos. Cada aspirante debía proporcionar su propio armamento y equipo inicial. Cada una de las curias de la república elegía a cien soldados de infantería y diez de caballería (López-Valdepérez, 2018).

Los incentivos principales eran la alimentación, un salario y, en ciertos casos, el producto del saqueo tras una victoria. A los soldados que cumplían su servicio militar sin sanciones se les permitía regresar a su casa y cuando eran merecedores recibían la categoría de veteranos y la honestas missio, una recompensa que consistía en dotación de tierras, dinero, exenciones de impuestos y el derecho de heredar la ciudadanía a sus sucesores; en estos casos estaban obligados a regresar al servicio activo cuando eran requeridos. Esta práctica se extendió a lo largo de toda la historia de la Roma antigua.



Marius sentado en una silla rodeado de oficiales romanos y bárbaros [Pintura], por Weston, W. & Rainey, P., 1990, Wikimedia Commons ([URL](#)).

Dominio público.



Cuando creció el poder de Roma, en principio, solo los ciudadanos tenían derecho a ingresar al ejército. En condiciones normales, el proceso de selección se conocía como *dilectus*, término que implica voluntad, honestidad y amor reflexivo y se volvió cada vez más riguroso. La primera selección dependía de una investigación de cada aspirante, debía ser ciudadano y se rechazaba a los delincuentes; se consideraban cartas de recomendación; se exigía una edad reglamentaria que para las fuerzas regulares era de 20 años; la estatura mínima iba de 165 a 172 cm, según su complejión y resistencia; se exigía saber leer y se daba preferencia a los más letrados. Al sistema de evaluación se le llamaba *probatio* y consistía en una serie de pruebas que incluían aptitud física, intelectual y moral, se daba importancia a los conocimientos jurídicos del aspirante, pero la prioridad era la disciplina y las habilidades específicas requeridas por el puesto a ocupar.

En estas condiciones, las pruebas eran de ejecución y de desempeño y las controlaban los oficiales. Existían dos procedimientos para seleccionar a los nuevos reclutas. El primer sistema era el *dilectus* (elección de los preferidos) que se realizaba en condiciones regulares, se llamaba a todos los ciudadanos que cumplían los requisitos legales y de entre ellos se elegía a los que pasarían a formar parte del ejército. El segundo sistema era el *tumultus*, que se realizaba bajo condiciones de necesidad urgente, como guerras o crisis políticas; en este caso se reclutaba a todos los candidatos disponibles y entonces se aceptaba incluso a quienes no cumplían los requisitos legales como los *proletarii* e aun se podía reclutar a esclavos.

Según Polibio (citado por López-Valdepéres, 2018) el nombramiento de los mandos era realizado primero en asambleas ciudadanas y más tarde por los órganos superiores de gobierno. Las tropas eran seleccionadas por los

oficiales asignando al azar al que elegiría a cada recluta, después de cada ronda se turnaban para que las legiones quedaran equilibradas en cuanto a sus efectivos. De esta manera, la selección se realizaba conforme al juicio de los oficiales.

En el periodo imperial (27 a.C. – 476 d.C.) Roma dependió fundamentalmente de su organización militar para conquistar territorios, controlarlos y administrarlos. De este modo, al mismo tiempo que se profesionalizó el ejército, creció en tamaño y complejidad organizativa y logística hasta alcanzar tal dimensión que fue indispensable incluir reclutas que no eran ciudadanos romanos. Al mismo tiempo, la diversidad de condiciones territoriales y culturales entre las partes del imperio exigió hacer más flexible el sistema de reclutamiento y selección. En tales condiciones fue necesario dar más importancia a la evaluación del desempeño de los soldados en servicio que a la selección para el ingreso. Así, la permanencia y ascenso dependía fundamentalmente del *cursum honorum*, la hoja de servicios de cada soldado.

En cuanto a los puestos en la administración pública, como ya se vio, en principio estaban reservados para los patricios. La selección dependía de la elección por los ciudadanos o bien, en el caso de designaciones, por los integrantes de los cuerpos colegiados de gobierno como el senado. Para el ascenso, el *cursum honorum* consideraba los cargos desempeñados registrados en el *curriculum vitae*.

De este modo, se observa que las evaluaciones más importantes en la Roma antigua estuvieron basadas en pruebas de ejecución y en el desempeño mostrado en el ejercicio de la tarea.



Edad Media

Con la caída del Imperio Romano de Occidente se produce una severa involución política en Europa. El impresionante desarrollo económico fue sustituido por el caos y la inseguridad, cesó el intercambio comercial alrededor de todo el Mediterráneo, terminó drásticamente la unidad administrativa y el territorio se desintegró en numerosos reinos cristianos y en no pocos casos se regresó a las ciudades-estado prerromanas. La fragmentación cultural fue inevitable y el latín, la lengua franca con la que se comunicaba el enorme territorio imperial, fue reemplazado por múltiples lenguas locales, aunque muchas de ellas, las lenguas romances, mantuvieron una clara influencia latina.

En esas condiciones era virtualmente imposible desarrollar un sistema de gobierno de gran alcance. La educación, por su parte, se limitaba a pequeños establecimientos privados en los que unos pocos que podían pagar aprendían religión y lo básico de lectura, escritura y aritmética. En algunos sitios funcionaban escuelas municipales y la educación más alta se impartía en escuelas monásticas y catedralicias; la nobleza y la clase gobernante se formaban en la Escuela de Palacio, lo que hoy llamamos aulas.

Así, la educación se reducía a la que impartían maestros particulares a los que contrataban los padres de familia para enseñar a sus hijos en casa a leer y escribir, principios morales y religiosos y, para aquellos que llegaban a estudios más avanzados, conocimientos teóricos que debían repetir de memoria ante el maestro o ante la familia. Por otra parte, las prácticas para seleccionar personal para los feudos o ciudades se reducían a las recomendaciones de quienes destacaban por su aptitud, competencia o preparación para tareas específicas. En general, no se encuentran registros de que hayan existido

sistemas de evaluación, pruebas o exámenes más o menos formales en este periodo temprano.

Uno de los reinos que surgió de la desintegración del Imperio Romano fue el de los francos, que en la primera mitad del siglo IX alcanzó su primera época de alto desarrollo con el Imperio Carolingio. Carlomagno se empeñó en reconstruir la unidad política en un gran reino cristiano y, como parte de su proyecto, en 781 convenció a Alcuino, un teólogo inglés, para conducir su Escuela de Palacio.

Las artes serviles fueron esenciales para el desarrollo técnico durante la Edad Media, realmente a ellas se deben los logros más importantes en esta época. Las artes serviles estaban formadas por los diversos oficios y trabajos técnicos que realizaban los siervos y los esclavos, tales como albañilería, herrería o carpintería. En los niveles profesionales, algunos de los oficios más importantes para la sociedad se organizaban en gremios o fraternidades y sus practicantes se formaban realizando las diversas tareas de cada oficio, ascendiendo desde los puestos elementales hasta las categorías de oficial y maestro. En este proceso, la evaluación de competencias la controlaba un oficial o maestro ya habilitado y dependía del desempeño que el aprendiz demostraba mediante pruebas de ejecución de las tareas; en los casos más exigentes, las pruebas se realizaban ante el gremio de maestros. Los conocimientos superiores del oficio eran celosamente guardados y se transmitían solo a los iniciados que se consideraban merecedores, dando origen a instituciones iniciáticas como es el caso de los



Universität Bologna Deutsche Nation [Imagen], XV, Wikipedia ([URL](#)). Dominio público.



masones, en el oficio de la albañilería. A estos gremios se debe la construcción de las catedrales.

En lo referente a la educación de orden más intelectual, en el contexto de las grandes diferencias entre las regiones de Europa, la educación quedó casi por completo bajo el control de la iglesia. Funcionaban las escuelas monásticas, donde se enseñaba a los niños, casi siempre de familias acomodadas, en las primeras letras y a los más avanzados en el trivium (gramática, dialéctica y retórica) y el quadrivium (aritmética, geometría, astronomía y música). En cuanto a la educación superior, el desarrollo de los studium generale en las escuelas catedralicias llevó al surgimiento de las primeras universidades a partir del siglo XI. Los programas educativos, así como los grados académicos y los exámenes, estaban reglamentados en cada universidad, aunque había diferencias entre instituciones.

En cuanto a la evaluación del aprendizaje, en la universidad medieval, en general, los estudiantes pagaban su matrícula y estudiaban en forma bastante libre los cursos, sin exámenes de promoción que condicionaran su permanencia. Pero no era así en el momento del egreso, veamos el caso del doctorado. Cuando su asesor (conciliarus) consideraba que un estudiante estaba suficientemente preparado, lo promovía para el examen privatum, que era el examen propiamente dicho. El asesor presentaba al estudiante al rector, ante quien debía jurar cumplir con los estatutos universitarios y no intentar corromper a los examinadores. Una semana antes del examen, el asesor se hacía responsable ante el arcediano (archidiácono, responsable de vigilar que los demás diáconos

cumplieran rectamente su oficio) de la capacidad del candidato para sustentar el examen. La mañana del examen, después de oír la misa del Espíritu Santo, el candidato se presentaba ante el claustro de doctores y se le daban dos textos para que los comentara; el candidato entonces se retiraba a preparar el comentario y por la tarde, en público, por lo común en la catedral, con la presencia del arcediano quien no podía intervenir, hacía su exposición ante un jurado de doctores; al terminar la exposición los doctores se retiraban a deliberar y, si se acordaba al menos mayoría de votos, el arcediano proclamaba el resultado y otorgaba al candidato el grado de licenciado (Le Goff, 1996).

Sin embargo, aun faltaba un segundo examen, el examen público (conventus publicum, doctoratus), para obtener el doctorado y tener derecho a enseñar magistralmente en cualquier universidad de la cristiandad. Conducido formalmente a la cátedra, el licenciado pronunciaba un discurso y presentaba una tesis que debía defender ante los estudiantes presentes, quienes tenían derecho a cuestionarlo sobre el tema de tesis; de esta manera demostraba su capacidad como maestro en una disputa universitaria. Aceptada su defensa de la tesis, era proclamado como doctor y el arcediano le entregaba la licencia para ejercer la cátedra, un libro abierto, un anillo de oro, la toga y el birrete.

Aquí se puede observar con bastante claridad el origen de las prácticas académicas que se siguen en las universidades actuales.

Al terminar la Edad Media, con la caída del Imperio Romano de Oriente y el inicio de la colonización de América, se redujo gradualmente el predominio de la iglesia y surgieron esfuerzos cada vez más extendidos para revivir la tradición cultural grecolatina y regresar al hombre al centro de interés. Se impulsó un avance creciente de la filosofía liberal y con los nuevos descubrimientos creció la necesidad de aumentar el conocimiento matemático y geográfico, así como el de las demás ramas de las nuevas ciencias, que entonces estaban surgiendo. Esto produjo uno

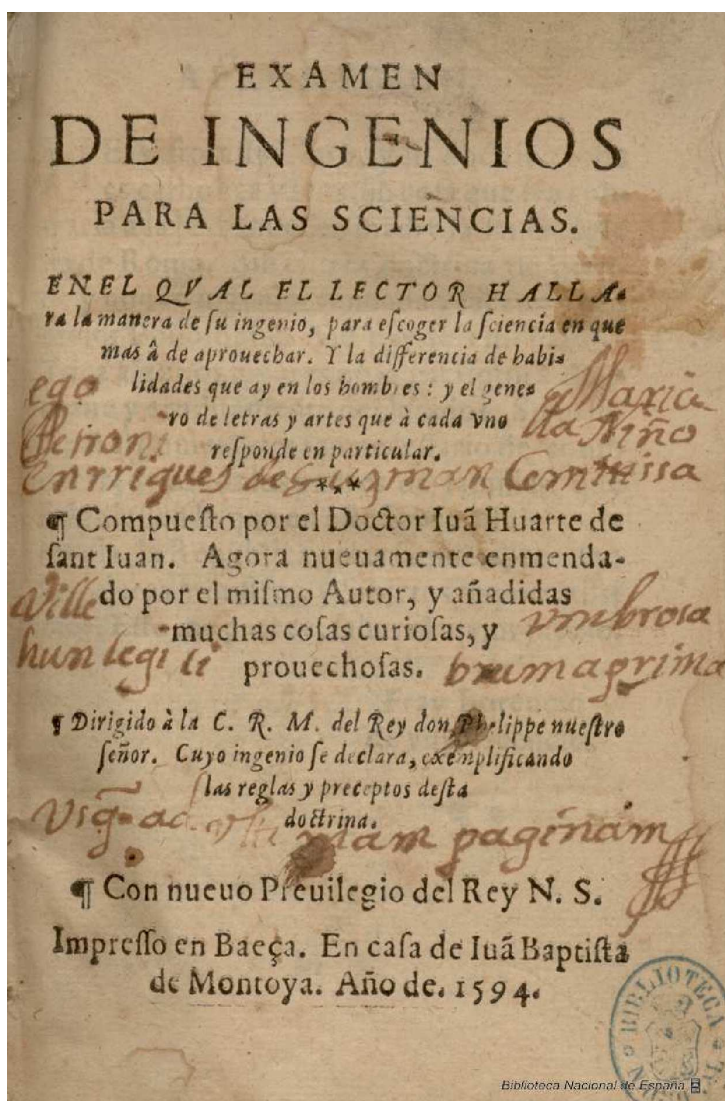
Edad Moderna

de los mayores cambios de la historia en la actitud hacia la cultura y el conocimiento.

El Renacimiento, un periodo que se extendió por más de un siglo, de 1400 a 1520, puede verse como la transición entre la Edad Media y la Edad Moderna. Si en el medievo, durante un milenio el aprendizaje, y consecuentemente su evaluación, tuvieron un carácter eminentemente memorístico y estuvieron dominados por la religión, en el Renacimiento se reaccionó a este estado de cosas avanzando gradualmente hacia un cambio de paradigma que hizo pasar el centro de atención de la teología al humanismo.

En lo que aquí interesa, la evaluación del aprendizaje se reorientó hacia la comprensión, la interpretación y la aplicación de conceptos, así como a la prevalencia del pensamiento crítico y analítico sobre la repetición puramente memorística. Este cambio, fundamental en la forma en que se concebía el aprendizaje, necesariamente afectó a su evaluación a través de exámenes. Es particularmente interesante que de las preguntas cerradas, que exigían respuestas únicas que dominaban en el medievo, en el Renacimiento comenzaron a promoverse las cuestiones abiertas, que aceptan respuestas diversas, pasibles de comprobación y que estimulan la creatividad.

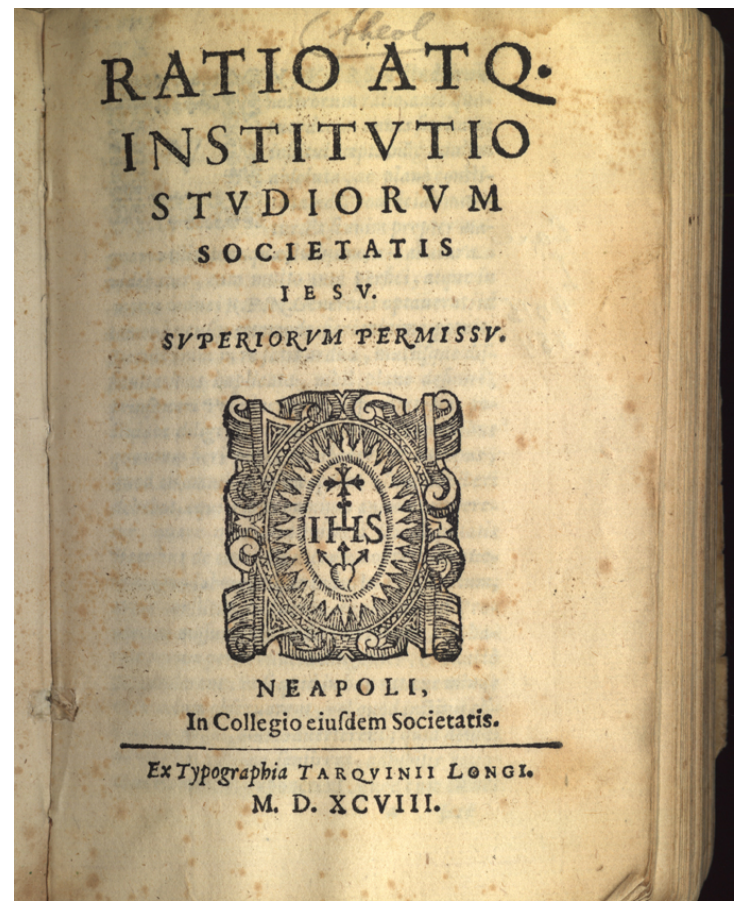
En 1575 el médico español Juan Huarte de San Juan (1529-1588) publica Examen de ingenios para las ciencias, obra en la que propone una teoría que sostiene que el ingenio y el entendimiento (inteligencia le llamamos hoy) de cada individuo no es de naturaleza espiritual ni depende de la voluntad, sino que obedece a causas biológicas, en particular a la constitución



Examen de ingenios para las ciencias [Fotografía], por Huarte de San Juan, J., 1603, Wikimedia Commons ([URL](#)) Dominio público.

del cerebro. A su vez, el ingenio y el entendimiento determinan la aptitud del sujeto para desempeñar cada ciencia y, en consecuencia, deben ser tomados en cuenta para decidir cuál es la mejor ocupación a la que debe dedicarse cada uno (Huarte, 1575/2003). Si bien Huarte, como era usual en su tiempo, se apoya en los filósofos de la Grecia clásica para proponer su teoría y explica el funcionamiento del cerebro con base en la teoría de los humores de Hipócrates y Galeno, su propuesta implica ya una concepción psicológica sobre la orientación vocacional y la selección de personal y plantea la necesidad de evaluar las diferentes aptitudes (clases de ingenios, les llama) necesarias para dedicarse a cada rama de la ciencia. En términos actuales, Huarte propuso medir las aptitudes del individuo para determinar la ocupación a la que debía dedicarse. Dado el contexto histórico, la obra de Huarte fue duramente censurada por la Inquisición y fue obligado a corregirla más de una vez. Sin embargo, tuvo una gran difusión y fue traducida a varios idiomas, influyendo sobre las ideas filosóficas que entonces había sobre la naturaleza humana, especialmente sobre la inteligencia y la aptitud.

A lo largo del Renacimiento la sociedad se independizó gradualmente del poder de la religión. Sin embargo, fue precisamente dentro de la iglesia donde surgió uno de los más importantes avances modernizadores de la educación en general, y en particular de la evaluación del aprendizaje, que tuvo un extenso impacto a lo largo del mundo. En un esfuerzo para contrarrestar el avance de la reforma luterana, en 1534 se funda la Compañía de Jesús, que pronto extendió sus misiones a América y Oriente. Aunque no estaba en su plan original, la Compañía comenzó a participar en la educación y, notablemente adaptada al marco del humanismo renacentista, impulsó profundos



Ratio Studiorum Societatis, 1598-1599 [imagen], por Chircho, 2005, Wikimedia Commons [URL](#). Dominio público.



Jan Amos Comenius (Komensky) (1592-1670) [Pintura], por Ovens, J., Wikimedia Commons [URL](#). Dominio público.

cambios en sus prácticas. En 1599 aprobó la Ratio atque Institutio Studiorum Societatis Jesus (Método y organización de estudios de la Compañía de Jesús) que constituye quizá el primer modelo curricular normado y regulado -hoy diríamos estandarizado- que se desarrolló durante más de un siglo, aplicándose a centenares de centros educativos jesuitas (Compañía de Jesús, 1599/2005). En lo relativo a los exámenes, tanto los destinados a los profesores como los que se aplicaban a los estudiantes, la Ratio establece reglas precisas -algunas sorprendentemente modernas- para su administración, calificación, interpretación y decisión (Soto, 1999). Como sistema, la Ratio

fue pionera en las prácticas de evaluación del aprendizaje.

En esta época, Juan Amós Comenio (1592-1670) fue uno de los iniciadores de la revolución educativa al promover la concepción de un sistema educativo formado por cuatro niveles, educación materna, primaria, secundaria y terciaria, que debía tener un carácter universal, incluyendo a las mujeres; el aprendizaje debía basarse en la experiencia, con un sentido práctico y holístico, también impulsó los libros de texto ilustrados para sistematizar y estandarizar la enseñanza y, consecuentemente, la evaluación del aprendizaje. Se promovieron los exámenes orales y públicos, realizados para demostrar los resultados obtenidos después de un ciclo de enseñanza. En 1630 publicó su *Didáctica Magna*, obra revolucionaria que se considera fundadora de la pedagogía moderna (Comenio, 1632/1998).

La obra educativa de los jesuitas contribuyó al advenimiento de la Ilustración durante el siglo XVIII, aunque no participase de algunos de sus principios fundamentales. La extensión y la influencia que para entonces había ganado la Compañía de Jesús fue tal que la hizo entrar en confrontación con el poder de la monarquía y, al negarse a aceptar los principios que esta le quería imponer, en 1762 se dio la ruptura definitiva cuando el Parlamento de París decretó la supresión de todas las escuelas de la Compañía de Jesús y el secuestro de sus bienes. Cinco años después, los jesuitas también fueron expulsados de los dominios de España.

Por otra parte, en el campo de la astronomía, desde antes de la invención del telescopio, el danés Tycho Brahe (1546–1601) encontró que cada medida observacional tiene un grado de error y propuso que para incrementar la precisión se pueden realizar varias mediciones y



Caricatura de dos jesuitas venecianos [Imagen], por Leone Ghezzi, P., 1752, Wikimedia Commons ([URL](#)). Dominio público.

obtener la media aritmética de ellas. Lo importante de la propuesta de Brahe es que utiliza, quizá por primera vez, las medidas estadísticas para controlar los errores perceptuales (Brahe, 1588/2010).

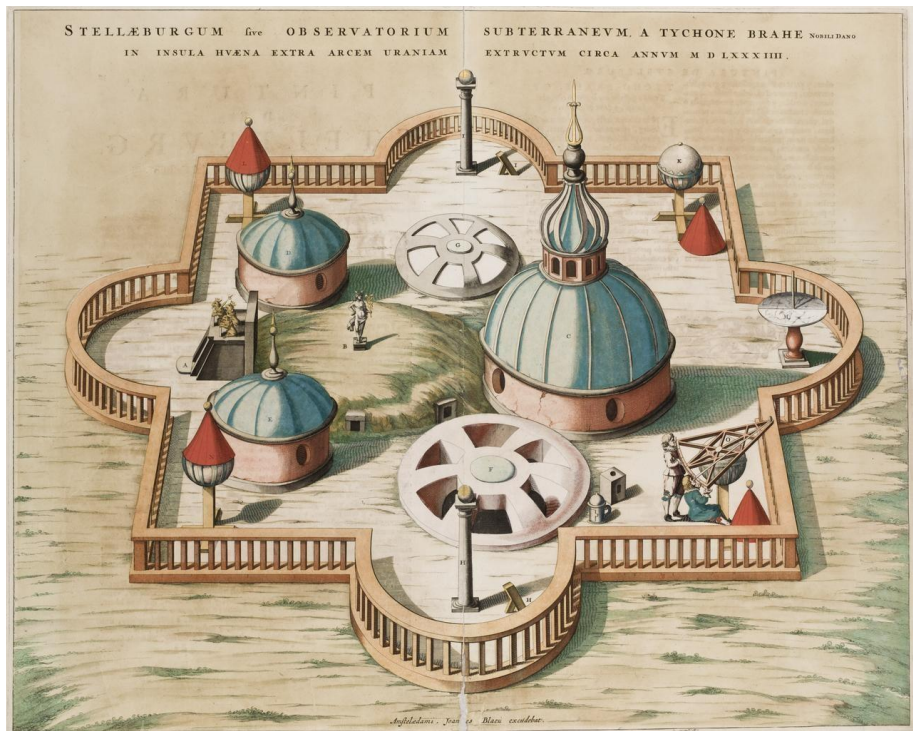
En el siglo XVII se produce la primera Revolución Científica. Inventos como el telescopio y el microscopio, el desarrollo del cálculo y la teoría de la probabilidad, así como descubrimientos y desarrollos científicos como la electricidad, las leyes de la gravitación y la presión atmosférica, por solo mencionar algunos, cambiaron radicalmente nuestro conocimiento y comprensión del mundo. Las lenguas vernáculas desplazaron al latín y la autoridad civil, monárquica en la mayoría de los casos, comenzó a intervenir en la educación para ampliarla.

En 1632 Galileo Galilei (1564-1642) publica su *Diálogo sobre los dos máximos sistemas del mundo, ptolemaico y copernicano*, libro fundamental para el nacimiento de la ciencia moderna, que le traería la persecución por la Inquisición (Galileo, 1632/2012). En esta obra, siguiendo las ideas de Copérnico, muestra cómo utilizando el telescopio (perfeccionado por él) y un compás, pudo medir el diámetro de la luna, la altura de sus montañas, la distancia de la tierra al sol y el movimiento de la tierra. Sin embargo, hace ver que las limitaciones de los sentidos, así como la instrumentación y su operación limitan la precisión de la medida.

En la década de los 1760 James Watt (1736–1819) diseñó una máquina de vapor con un condensador separado, innovación tecnológica

que mejoró su eficiencia y la hizo viable como fuente de energía, que a su vez permitió el desarrollo de máquinas para fábricas, minas, la agricultura y el transporte. Así dio inicio la primera Revolución Industrial.

En 1786, Immanuel Kant (1724-1804) publicó *Primeros principios metafísicos de la ciencia de la naturaleza*, obra en la que busca conocer los principios a priori que hacen posible el conocimiento. En el Prefacio afirma que la psicología empírica no puede ser una ciencia de la naturaleza, porque la matemática no puede aplicarse a los fenómenos del sentido interno; además, al someter a otro sujeto pensante a



*Sjernerborg Observatory [Impresión], por Blaeu, J., 1634, Jstor ([URL](#)).
CC BY NC SA 4.0.*

observación, ésta por sí misma altera y distorsiona el estado del objeto observado (1786/1993). Así, desde la filosofía, se declaraba la imposibilidad de conocer científicamente los fenómenos psicológicos.



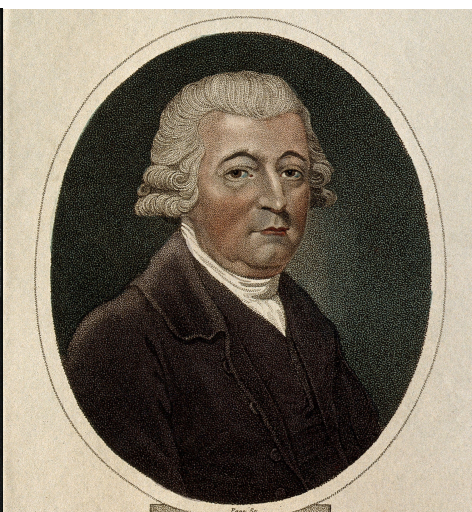
Siglo XIX

Este es el siglo en el que se sitúa el origen de la psicología científica, el cual está asociado con el desarrollo de métodos de medición experimental y controlada de los fenómenos psicológicos. Sin duda, la Revolución Industrial tuvo un impacto decisivo para el desarrollo de instrumentos de alta precisión para medir variables de naturaleza puramente física como el tiempo, la masa, la distancia, el sonido o la temperatura, lo que abrió nuevas posibilidades para más adelante medir los tiempos de reacción. Como referencia, Alexander von Humboldt (1769-1859) fundó el primer laboratorio científico en 1810 en la Universidad de Berlín.

En 1823, Friedrich Wilhelm Bessel (1784-1846) publicó Sobre la corrección de los errores personales en las observaciones astronómicas, artículo en el que reporta que al analizar los tiempos registrados por distintos astrónomos sobre el tránsito de Venus, había diferencias significativas. Al encontrar que tales errores tenían un carácter sistemático, propuso que estas diferencias se debían a factores



Il Divino Galileo Divincenzio Galilei
[Grabado], por Allegri, F., & Zocchi, G.,
Istor ([URL](#)). Dominio público.



Nevil maskelyne [Grabado], Page, R., 1815,
Wellcome Collection ([URL](#)). Dominio público.



Friedrich Wilhelm Bessel [impresión], Wolff, J.
& Mandel, E., Wellcome Collection ([URL](#)).
Dominio público.

En 1795 ocurre un suceso que pasaría a la historia como uno de los precursores de la medición de los fenómenos psicológicos. Nevil Maskelyne (1732-1811), director del Real Observatorio de Greenwich, consideró necesario despedir a su ayudante David Kinnebrook porque comprobó que en sus mediciones cometió errores mayores a los que él mismo cometía (Maskelyne, 1795). Nunca pudo explicar la razón de estos errores.

individuales que afectan al tiempo de reacción, tales como la atención, el juicio o la preferencia, y que podían expresarse mediante una ecuación que permite calcular el error medio de cada observador (Bessel, 1823). Esta pasaría a la historia como la Ecuación personal, quizá la primera formulación matemática, producto de observaciones empíricas y críticas, que trata de relacionar factores humanos, de carácter individual, que dificultan la observación objetiva y exacta, así como el razonamiento único.

En 1838, Bessel propuso que las diferencias individuales en la percepción y el procesamiento de información de cada observador pueden determinarse cuantitativamente. Si definimos a t como el tiempo medido por el observador; T como el tiempo verdadero del evento astronómico; C como la constante personal del observador (un valor signado) y; E como el error de medición; entonces $t=T+C+E$. El valor de C puede determinarse comparando la medida del observador contra las medidas de otros observadores (será negativo cuando es menor a la media y positivo cuando es mayor); cuando se estabiliza por un número suficiente de medidas, se convierte efectivamente en la constante del observador analizado (Bessel, 1875). De esta forma, la ecuación personal cuantifica el factor humano.

En este punto conviene observar que actualmente, desde la pseudociencia, se ha utilizado el término “ecuación personal” con sentidos fantasiosos, que nada tienen que ver con el estudio de los factores perceptuales.

Tal era el estado del arte cuando, apenas una década después de las observaciones de Bessel, en 1834 Ernst Heinrich Weber (1795–1878), médico alemán, publica en Leipzig De

tactu, obra en la que reporta sus experimentos sobre la medición sistemática y controlada de la sensación por medios experimentales; utilizando instrumentos elementales como termómetros, compás e incluso monedas, logra establecer, por primera vez, una relación cuantitativa entre la magnitud del estímulo y su percepción por el sujeto (Weber, 1834). Esta es una obra central en la historia de la psicología.

En 1850, en un experimento clásico, el médico y físico Hermann Ludwig Ferdinand von Helmholtz (c. 1821–1894), con el fin de medir el tiempo que tomaba a la contracción muscular, aplicó un estímulo a un nervio motor en un punto cercano a la unión neuromuscular y, bajo otra condición, en un punto mucho más alejado. En otras palabras, midió la velocidad de la transmisión nerviosa sin intermediación del cerebro, encontrando que no era mayor a 100 pies por segundo. Los pájaros vuelan más rápido, los caballos y el movimiento de nuestro brazo casi alcanzan esta velocidad. Resultados similares se han obtenido con ranas. En animales de sangre caliente, en humanos en particular, Helmholtz excitó puntos de la piel a diferentes distancias del cerebro y midió el tiempo de respuesta de un signo determinado, por ejemplo un movimiento de la mano; en humanos la velocidad fue de 200 pies por segundo, el doble que en nervios de ranas (Helmholtz, 1903). Sin embargo, este método tiene varias limitaciones y sus repeticiones obtienen resultados divergentes; con métodos más elaborados, posteriormente se encontró que los resultados en humanos son similares, apenas poco más rápidos, a los que se obtienen en ranas. La velocidad de conducción en los nervios sensoriales es la misma que en los motores.

En 1860, Gustav Theodor Fechner (1801-1887), filósofo y psicólogo alemán, publica *Elements der Psychophysic*, obra que se reconoce como el origen de la psicología experimental. A partir



View of Flamsteed House within the courtyard-Royal Observatory Greenwich Ca 1900 [Fotografía], Flickr ([URL](#)). Dominio público.



Ernst Heinrich Weber, 1795-1878, head-and-shoulders portrait, facing front [Litografía], por Hoffmann, R., 1856, Library of congress ([URL](#)). Dominio público.

Hermann Ludwig Ferdinand von Helmholtz [Fotografía], por Stacey, WJ., Wellcome Collection ([URL](#)). Dominio público.

Portrait of Gustav Theodor Fechner (1801-1887) [Fotografía], Flickr ([URL](#)). Dominio público.

Dr. Adolphe Hirsch [fotografía], en Bulletin de la Soci neuchloise des sciences naturelles (p. 453), 1898, Archive ([URL](#)). Dominio público.

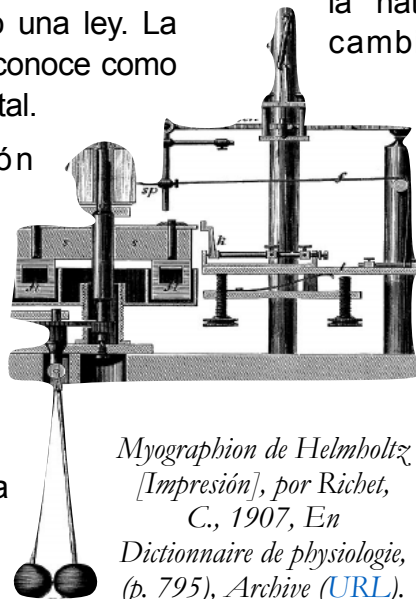
Franciscus Cornelius Donders, + am 24. März [Grabado], het utrecht archief ([URL](#)). Dominio público.

de los experimentos de Weber, Fechner desarrolla los métodos de estímulos constantes, de las diferencias apenas perceptibles y del error medio y logra formalizar matemáticamente la relación entre la intensidad del estímulo físico y los cambios en la sensación, conocida como la Ley de Weber-Fechner (Fechner, 1860). Aunque Ernest Weber había descubierto esta relación en 1834, no llegó a formalizarla como una ley. La publicación de los Elementos se reconoce como el origen de la psicología experimental.

De nuevo en la observación astronómica, inspirado por la ecuación personal y los experimentos de Helmholtz, Adolphe Hirsch (1830–1901), astrónomo nacido en Alemania, utilizando el cronoscopio midió controladamente el tiempo transcurrido entre el estímulo y la respuesta, tanto en el tacto al estimular la piel en puntos a distintas distancias del cerebro, como en el oído y el ojo (Hirsch, 1864).

En 1868, una década antes del Laboratorio de Wundt, el fisiólogo y oftalmólogo Franciscus Cornelius Donders (1818–1889), reaccionando al tratamiento abstracto y especulativo que la

filosofía seguía dando a los fenómenos mentales y siguiendo los planteamientos sobre la ecuación personal, así como los experimentos sobre la sensación de Weber y Fechner, al preguntarse sobre lo que ocurre en el cerebro durante los procesos de sensación, razonamiento y volición, consideró poco probable que alguna vez pudiera alcanzarse una respuesta concluyente sobre la naturaleza de estos fenómenos. En cambio, decidió medir bajo control experimental riguroso y con los instrumentos más avanzados de su época (algunos diseñados por él mismo) el tiempo que toman los procesos mentales desde la aplicación del estímulo hasta la emisión de la respuesta; a éste se le llamó tiempo de reacción y desde entonces constituye una variable dependiente central en el estudio de la psicofísica. Por ejemplo, en un experimento con un balancín se aplicaban choques eléctricos en uno u otro pie; había dos condiciones, en la primera el sujeto sabía el pie que sería estimulado, en la segunda no; en ambas condiciones el sujeto debía responder con la mano del lado estimulado; se encontró que el

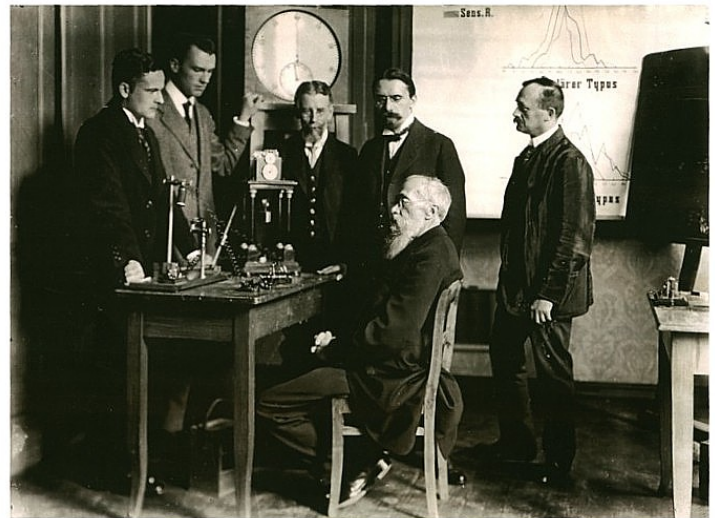


Myographion de Helmholtz [Impresión], por Richet, C., 1907, En Dictionnaire de physiologie, (p. 795), Archive ([URL](#)). Dominio público.

tiempo de reacción era mayor bajo la segunda condición. Esta fue la primera vez que se midió el tiempo de reacción en un proceso nervioso bien definido. Los experimentos se replicaron en vista y oído. Debe observarse que el tiempo transcurrido entre la aplicación del estímulo y la emisión de la respuesta, cualquiera que sea el órgano sensorial involucrado, incluye los procesos mentales mediadores (Donders, 1969).

En 1879, Wilhelm Maximilian Wundt (1832-1920), médico, psicólogo y filósofo alemán, fundó el primer laboratorio de psicología experimental en Leipzig, de donde saldrían los primeros psicólogos formados académicamente. Ahí se impulsó la investigación científica de fenómenos psicológicos fundamentales como la percepción, la memoria, la atención y la emoción, utilizando la introspección como método fundamental. Se desarrollaron métodos experimentales que se basaban en la medición psicofísica, para lo cual se utilizaron los instrumentos más avanzados de que se disponía en su época y se desarrollaron otros (Wundt, 1897).

Francis Galton (1822-1911), polímata inglés muy influido por la teoría de la evolución, trató de explicar la inteligencia mediante factores hereditarios y promovió la psicología diferencial y la eugenesia. Instaló el Laboratorio Antropométrico en un pabellón de la Exposición Internacional de Salud de 1884 en Londres, donde midió sistemática y controladamente una serie de variables utilizando aparatos de precisión y hojas de registro para los datos; tanto los instrumentos como los sistemas de registro y procesamiento de datos en no pocos casos fueron diseñados por el propio Galton y sirvieron de prototipo para desarrollos posteriores. Se medía la estatura de pie y sentado, peso, envergadura (amplitud total con los brazos extendidos), potencia respiratoria, fuerza de tracción y de apriete, fuerza al soplar, agudeza auditiva y visual, sensibilidad al color y otras variables. Estas mediciones buscaban dos



*Wundt research group, ca. 1880 [Fotografía],
Wikimedia Commons ([URL](#)). Dominio público.*

objetivos: en el nivel personal, la información debía servir para diagnosticar el desarrollo y propiciar la intervención oportuna y, en el nivel estadístico, para conocer y seguir los cambios de la eficiencia alcanzada en grupos tales como escuelas, barrios, poblaciones, ocupaciones o razas (Galton, 1884). En estadística, aplicó la distribución normal y formuló el análisis de correlación para los estudios en psicología diferencial. Sostuvo que las habilidades psicológicas son producto de la herencia en la misma medida que lo son las características físicas del individuo. En 1904 fundó la Oficina de Registro de Eugenesia, que en 1907 se convirtió en el Laboratorio de Eugenesia Galton dentro del University College de Londres, bajo la dirección de Karl Pearson, y en 1996 pasó a formar parte del Departamento de Biología del University College.

Hermann Ebbinghaus (1850-1909), psicólogo alemán, publicó *Sobre la memoria* (Ebbinghaus, 1885), obra en la que reporta sus estudios experimentales y formalización matemática sobre la memoria, la curva del olvido y el efecto de espaciado, en relación con el aprendizaje. Inspirado en el método de Fechner, se usó a sí

mismo como sujeto para experimentar sobre la capacidad de reproducción de sílabas sin sentido, formadas por una vocal entre dos consonantes. Este es el primer intento de medir experimentalmente la memoria (y probablemente cualquier otro proceso psicológico superior), algo a lo que ni Weber, ni Fechner ni Wundt se atrevieron.

James McKeen Cattell (1860–1944), psicólogo estadounidense, se doctoró y trabajó con Wilhelm Wundt en Leipzig y colaboró en las investigaciones de Francis Galton; ocupó las más altas posiciones dentro de la psicología americana. Congruente con el espíritu de su tiempo, Cattell se interesó más en el método que en la teoría, así como en asegurar la precisión en el conocimiento de los hechos mediante la mejora continua de instrumentos para presentar estímulos y registrar respuestas. Trabajó en los tiempos de exposición para leer, reconocer palabras, colores y objetos y en los tiempos de asociación. Reinterpretó la Ley de Weber-Fechner, proponiendo que la precisión aumenta al interpretar las pequeñas diferencias mediante la raíz cuadrada, conocida como la Ley de Fullerton-Cattell (Fullerton&Cattell, 1892). En la Universidad de Columbia, aplicó tests a 100 estudiantes en 1896, una década antes de Binet; a diferencia de éste, Cattell interpretó los puntajes por su posición relativa (Cattell, 1896).

Puede observarse que al terminar el siglo XIX la medición de los fenómenos psicológicos se orientaba, por un lado, a estudiar los tiempos de reacción, aplicando estímulos y midiendo el tiempo que tardaba en emitirse la respuesta, todo esto en condiciones experimentales controladas y apoyándose en instrumentos de alta precisión, casi siempre diseñados especialmente; esta orientación surgió de las observaciones sobre los errores de medición en astronomía; los experimentos de Ernst Weber para medir la sensación, al ser continuados por

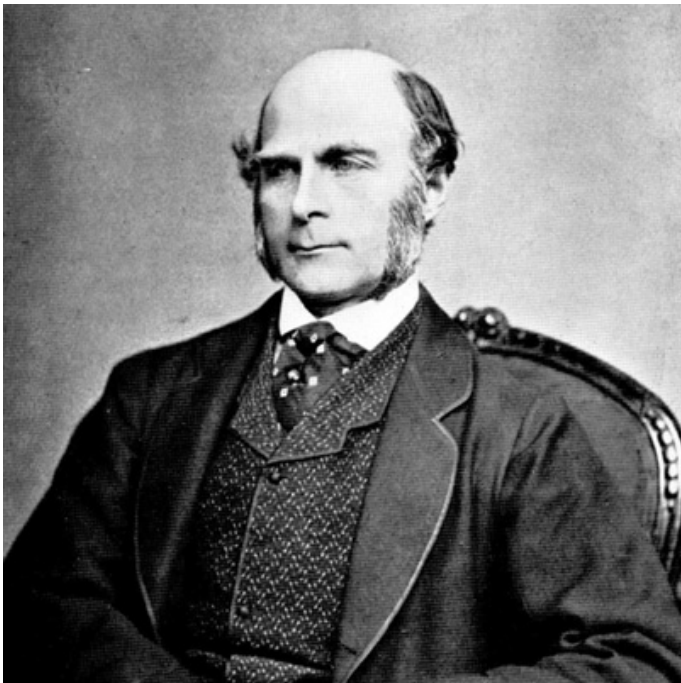
Theodor Fechner, dieron origen a la psicofísica y, en sentido estricto, a la psicología experimental. Por otro lado, siguiendo una orientación bastante diferente, Francis Galton intentó explicar los fenómenos psicológicos, principalmente la inteligencia, mediante la teoría de la evolución, la herencia y la biometría. Tales son los orígenes de la psicometría como hoy se le entiende.



Hipp chronoscope with small galvanometer and key [Fotografía], Hipp, M., 1888, Jstor ([URL](#)). CC BY NC SA 4.0.



El siglo XX es la época de mayor expansión de la psicometría. Los inicios del siglo presencian el cambio desde los estudios psicofísicos, caracterizados por la aplicación de estímulos físicos y las consecuentes respuestas consistentes en conductas abiertas, generalmente verbales o motoras, todo ello rigurosamente medido y registrado bajo condiciones controladas donde la precisión y objetividad eran la preocupación principal, para pasar al paradigma de los instrumentos que miden rasgos latentes de complejidad creciente; se podría decir que se pasó del paradigma entradas-salidas al que intenta estudiar el interior de la “caja negra”. En el segundo paradigma, que con frecuencia se identifica simple y llanamente como la psicometría, dominan los tests, instrumentos que presentan tareas de diferentes tipos a los que técnicamente se les llama “reactivos”, pues estrictamente son estímulos que se presentan al sujeto para provocar una reacción, que siempre consiste en algún tipo de conducta directamente observable.



Francis Galton 1850 [Fotografía], Wikimedia Commons ([URL](#)).
Dominio público.

Siglo XX

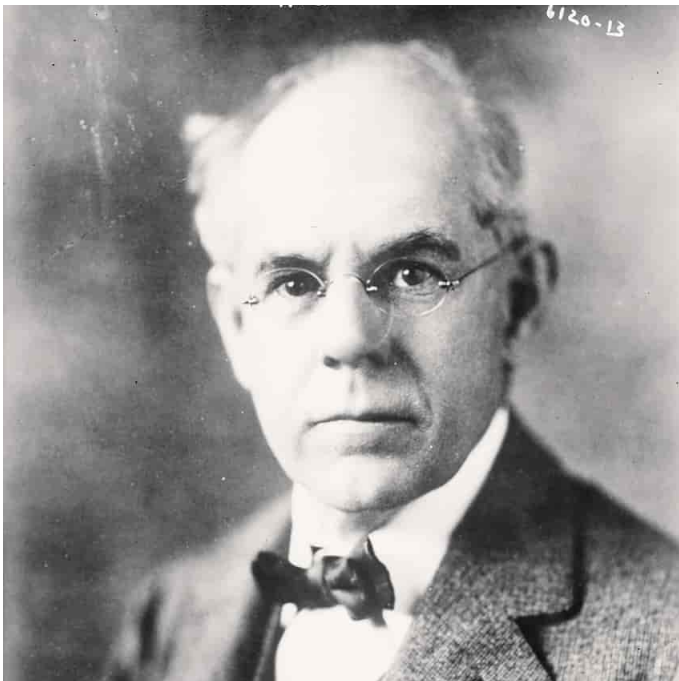
De este modo, pueden identificarse tres periodos en el desarrollo de la medición de los fenómenos psicológicos. El primero va desde finales del siglo XIX, en que se desarrollan las medidas psicofísicas y antropométricas y se producen los primeros intentos de medir atributos no observables directamente (rasgos latentes, se les llamará más tarde) como la inteligencia. En el segundo periodo, fuertemente influido y estimulado por las dos Guerras Mundiales, dominará la Teoría Clásica del Test (TCT), que busca crear instrumentos capaces de medir cada vez con mayor precisión los constructos que constituyen su objeto de estudio, convirtiendo a la confiabilidad y la validez en sus preocupaciones centrales, lo que hace que toda medida quede vinculada a una población de referencia para su interpretación y uso; más tarde, como reacción a la TCT, se propondrá la Teoría de la Respuesta al Ítem (TRI) que provocará un cambio radical de paradigma y cuestionará el concepto mismo de test. En el tercer periodo, que inicia después de la Segunda Guerra Mundial, como ocurrió con todo lo demás, la computadora tendrá una importancia central para la psicología.

Pasemos entonces a revisar algunos de los principales eventos que marcan el desarrollo de la psicometría en el siglo XX.

James McKeen Cattell (1860-1944), psicólogo norteamericano, trabajó en el laboratorio de Wundt y se doctoró bajo su dirección; poco después colaboró con Galton en la fundación del Laboratorio Antropométrico. Sin embargo, al principio Cattell trabajó más en los tiempos de reacción, utilizando diversos estímulos como

caracteres alfabéticos, colores e imágenes y variando sistemáticamente su administración; por ejemplo, presentando letras aisladas, palabras u oraciones con o sin sentido, en mayúsculas o minúsculas, con tipos góticos o latinos; para ello utilizó kimógrafos, cronómetros y otros instrumentos de alta precisión, como medir tiempo a nivel de milisegundos; la respuesta en general era verbal, por ejemplo leer en voz alta (Cattell, 1886). Más tarde, incursionó en la medición de constructos como la personalidad.

Cattell y otros, entre los que destaca Granville Stanley Hall (1844-1924), de Leipzig regresaron a Estados Unidos y apenas cuatro años después del laboratorio de Wundt, en la Universidad Johns Hopkins se fundó el segundo laboratorio de psicología; a partir de ahí comenzaron a abrirse laboratorios en las universidades norteamericanas. Las características principales de estos laboratorios eran tres: primera, se equipaban con instrumentos de medición especializados y de frontera para la aplicación



James McKeen Cattell [Fotografía], Store norske leksikon ([URL](#)). Dominio público.

controlada de estímulos, así como la medición objetiva y el registro de las respuestas; segunda, estaban orientados a aplicaciones educativas, clínicas e industriales; la tercera característica es que en ellos trabajaban estudiantes como parte del plan de estudios de psicología (Cattell, 1928).

La medición del aprendizaje ha sido una de las vías de desarrollo centrales de la psicometría, en particular la evaluación con fines de ingreso a la universidad y el control de la práctica profesional. Veamos algunos ejemplos representativos. En 1900 se fundó en Estados Unidos el College Board con el objetivo de apoyar a los estudiantes en la transición a la educación superior; actualmente tiene más de 6,000 instituciones asociadas, nacionales e internacionales; en 1901 aplicó el primer examen a 973 estudiantes de 67 localidades de Estados Unidos y dos de Europa; en 1926 aplicó por primera vez el SAT (Scholastic Aptitude Test). En cuanto a la práctica profesional, la necesidad urgente de elevar la seguridad de las construcciones en Estados Unidos, llevó a que desde fines del siglo XIX se legislara sobre el control del ejercicio de la arquitectura; en 1919 se creó el NCARB (National Council of Architectural Registration Boards) para certificar y expedir la licencia profesional; en la década de los 1980 se comenzó a aplicar el ARE (Architect Registration Examination); probablemente este es el primer test de habilitación profesional. Por otra parte, en 1949 el Educational Testing Service creó el GRE (Graduate Record Examinations) que se utiliza para la admisión al posgrado en general. En 1954 el Graduate Management Admission Council creó el GMAT (Graduate Management Admission Test), un examen de admisión para el posgrado en Administración. En 1964 se creó el TOEFL (Test of English as a Foreign Language) para evaluar el dominio del inglés por hablantes no nativos.



En las dos últimas décadas del siglo XIX, durante la Tercera República, Francia volvió legalmente obligatoria la instrucción básica para niños de 6 a 13 años. En consecuencia, aunque el proceso de enseñanza-aprendizaje podía realizarse en la familia o con profesores particulares, se impulsó la escuela pública laica y obligatoria como un elemento central del proyecto del estado-nación. En este contexto, el gobierno encargó al psicólogo Alfred Binet (1857-1911) y al psiquiatra Théodore Simon (1872-1961) desarrollar una forma de medir la inteligencia de los niños escolares, con el propósito de detectar retraso en el desarrollo y poder aplicar acciones remediales. En 1905 fue aplicado por primera vez el test de Binet-Simon en las escuelas públicas de París. Este fue el primer instrumento construido en función de tareas que involucran operaciones mentales como medio para medir inteligencia, la cual a su vez se utiliza como predictor del rendimiento escolar. Lo que provocó una ruptura respecto a los métodos anteriores, fue que este test presenta al niño tareas cognitivas que involucran la capacidad de lenguaje, memoria, juicio, razonamiento y comprensión. A partir de los

resultados de la aplicación del test, se inició el desarrollo de programas de intervención educativa para atender a los niños con desventaja. El test originalmente se formaba por veinte reactivos en escala de complejidad creciente. Se aplicaba a niños desde los tres hasta los quince años y las tareas iban desde decir su apellido y repetir tres cifras, hasta resolver rompecabezas y problemas visoespaciales. Aplicado a grupos etarios y procesado estadísticamente, definió operacionalmente el concepto de edad mental: se toma como puntaje de desempeño el nivel máximo de complejidad alcanzado por cada sujeto de un grupo etario, de manera que la media aritmética constituye esa edad mental; en otras palabras, la edad mental es el promedio de capacidad intelectual de un grupo etario. De este modo, una edad mental inferior a la edad cronológica implica retardo en el desarrollo, mientras que lo contrario implica dotación superior (Binet, 1904). Sobre esta base, poco después se desarrolló el concepto de cociente intelectual.

En 1908, en Estados Unidos, el psicólogo Henry Herbert Goddard (1866-1957) tradujo el test de Binet-Simon y lo aplicó a más de 20,000 niños escolares (Goddard, 1908). Defendió el uso de pruebas, propuso una clasificación en función de los puntajes obtenidos y promovió su uso con fines educativos, migratorios, legales y otros. Goddard fue un pionero en la aplicación masiva y sistemática de tests a grandes poblaciones con fines de clasificación (Goddard, 1913). Conforme al espíritu de la época, Goddard estuvo interesado en los estudios sobre herencia y eugenesia.

En 1916, el psicólogo alemán William Lewis Stern (1871-1938) formuló el concepto de cociente intelectual (intelligenzquotient) (Stern, 1916). En el nivel de diagnóstico individual, dividiendo la edad mental alcanzada



*William Stern (1871-1938) [Fotografía],
Wikimedia Commons ([URL](#)). Dominio público.*

por un sujeto entre su edad cronológica se obtiene el cociente intelectual; para interpretarlo se multiplica por cien y se redondea a enteros. De este modo, un CI = 100 indica que el sujeto tiene justo la inteligencia que corresponde a su edad; un CI < 100 indica deficiencia y un CI > 100 indica dotación.

En 1916, Lewis Madison Terman, pedagogo y psicólogo estadounidense, realizó la revisión conocida como Stanford-Binet, lo editó y promovió una aplicación aún más extensa que la de Goddard (Terman, 1917). Durante la Primera Guerra Mundial sirvió en el ejército aplicando, con un equipo de psicólogos, pruebas a los reclutas; la aplicación se estandarizó a una hora y conforme a los puntajes los resultados se clasificaban en cinco normas, de "A" a "E", y se decidía su asignación; la categoría "A" pasaba a la escuela de oficiales mientras que los puntajes "D" o "E" perdían la posibilidad de acceder al entrenamiento para oficial. A diferencia de Binet, Terman proponía utilizar el CI para diagnosticar y

canalizar al sujeto a la ocupación más adecuada conforme a su aptitud. En 1921 inició un programa para detectar, atender y dar seguimiento a los dotados, que se extendió hasta después de su muerte (Terman et al., 1917; 1925). El test Stanford-Binet se revisó en diferentes ocasiones, la última revisión se realizó en 2003 y actualmente aún se aplica para medir la inteligencia general en niños y adultos.

En 1917, Robert Mearns Yerkes (1876-1956) psicólogo estadounidense, como presidente de la APA impulsó diversos programas de psicología para apoyar el esfuerzo de guerra. Dirigió el comité de exámenes psicológicos para el reclutamiento donde, a partir del Binet-Simon, desarrolló, primero, el test Army Alpha para medir habilidad verbal, conocimientos e información y la capacidad para seguir instrucciones. Ante el hecho de que muchos soldados, más aún los inmigrados, no estaban alfabetizados, desarrolló después el Army Beta, una forma no verbal que utiliza un lenguaje icónico y diversos símbolos presentando tareas como completamiento de figuras, solución de laberintos y reconocimiento de patrones (Yoakum & Yerkes, 1920). Yerkes, colaborando con Terman, Goddard y otros psicólogos, desarrollaron métodos rápidos para la medición de constructos sobre las habilidades mentales.

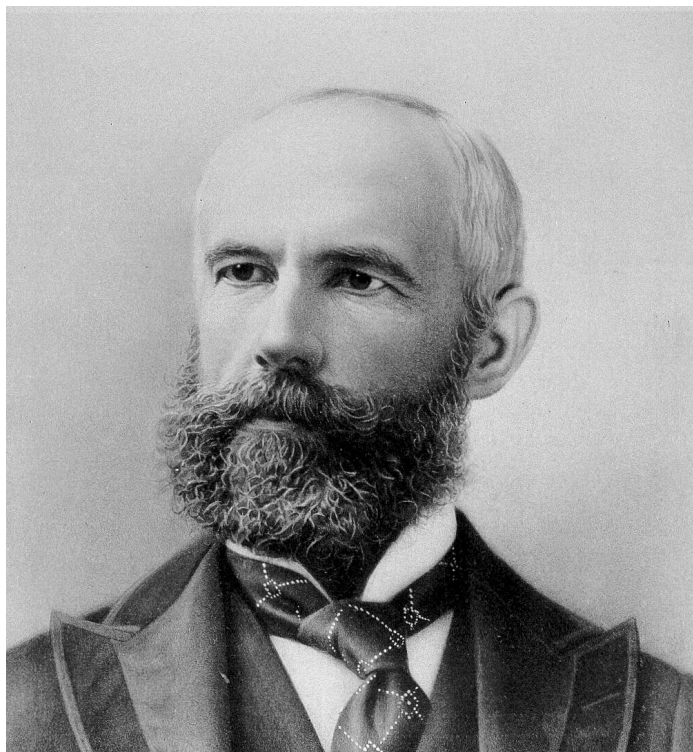
Charles Spearman (1863-1945), psicólogo inglés, formuló la teoría bifactorial de la inteligencia, que propone que la inteligencia se compone de dos factores, el factor general (G) de carácter hereditario y el factor especial (S) producto del desarrollo de cada sujeto. Ante la carencia de herramientas matemáticas adecuadas a sus propósitos, Spearman desarrolló el análisis factorial que, mediante correlaciones entre variables de nivel ordinal (coeficiente rho de Spearman), permite evaluar la capacidad explicativa (rendimiento) de cada variable; a partir de ahí propuso el modelo lineal,



Hermann Ebbinghaus (1850-1909) [Fotografía], Wikimedia Commons ([URL](#)). Dominio público.

que asume que el resultado de aplicar un test a una persona se expresa como un puntaje de carácter empírico que se representa por X y está compuesto por el valor verdadero (V) más un error (e) que siempre estará presente, incluso en medidas físicas (Spearman, 1927). Este se ha considerado como el inicio de la TCT.

Hermann Rorschach (1884-1922), médico psiquiatra y psicoanalista suizo que estudió con Bleuler y fue influido por Jung, en 1921 publicó el instrumento que lleva su nombre, formado por diez láminas con manchas de tinta que carecen de una configuración predeterminada (Rorschach, 1998). Este fue el primer test proyectivo publicado. Se trata de un instrumento psicométrico singular, porque solo tiene sentido como parte de una técnica aplicada por un psicoterapeuta experto dentro del proceso de análisis. Las láminas se presentan al paciente, quien debe decir qué ve o con qué asocia la imagen; sus respuestas no producen puntajes



G. Stanley Hall [Fotografiado], Wellcome Collection ([URL](#)). Dominio público.

sino que, como ocurre en los posteriores tests proyectivos, deben ser directamente interpretadas por el analista.

Robert Sessions Woodworth (1869-1962), psicólogo estadounidense, en 1917, en la Primera Guerra Mundial, desarrolló la “hoja de datos personales” (the personal data sheet); se formaba por 116 reactivos del tipo falso-verdadero y fue construida como una lista de síntomas a partir de los cuales se detectaba psiconeurosis; este es quizá el primer instrumento para medir personalidad (Woodworth, 1930). La Hoja no fue terminada a tiempo para su aplicación en el ejército, pero a partir de este instrumento precursor, el propio Woodworth y otros investigadores relevantes construyeron y analizaron miles de reactivos, hasta convergir en el modelo de rasgos conocido como los Cinco Grandes (apertura, responsabilidad, extraversión, amabilidad y neuroticismo) el cual, aunque ha sido fuertemente cuestionado, se sigue utilizando e investigando en la psicometría de la personalidad.

En 1935 se funda la Sociedad Psicométrica en Ann Arbor, Michigan. Contó entre sus fundadores con Albert K. Kurtz, Paul Horst, L. L. Thurstone, entre otros y fueron presidentes Thorndike, Guilford y Gulliksen, entre otros. Esta organización edita desde 1936 la revista Psychometrika, en la que se han publicado hasta la actualidad algunos de los artículos centrales en el desarrollo de la psicometría.

En 1935 se publicó el Test de Apercepción Temática (TAT) por Henry Alexander Murray (1893-1988), psicólogo estadounidense, y Christiana Drummond Morgan (1897-1967), artista estadounidense. El TAT es un instrumento proyectivo formado por 31 láminas que contienen imágenes estructuradas. Excepto la lámina 20, que presenta una figura difusa; las demás muestran imágenes claras, pero todas

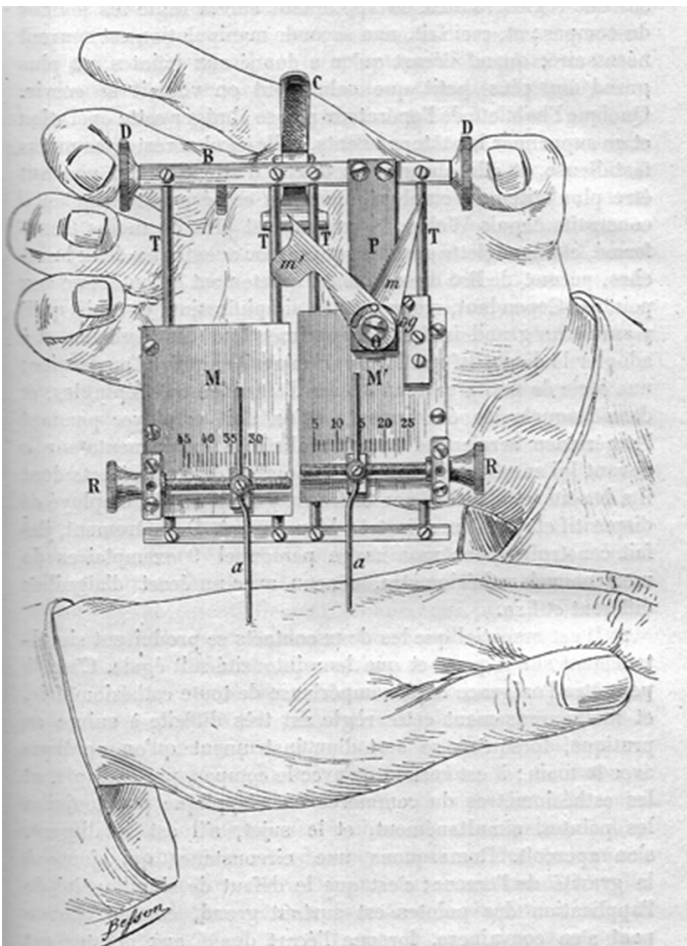
tienen un significado ambiguo y se refieren a diferentes temas. Once láminas son de aplicación universal y el resto se aplican según sexo y edad, de modo que a un sujeto se le pueden aplicar hasta veinte reactivos (Morgan, 1995).

Louis Leon Thurstone (1887-1955), ingeniero estadounidense que llegó a ser ayudante de Thomas Alva Edison, posteriormente se doctoró en psicología. Como parte de sus contribuciones matemáticas, aplicó la media y la desviación estándar para estandarizar los puntajes del CI y contribuyó al desarrollo del análisis factorial con técnicas que fueron aplicadas al desarrollo del WAIS e incluso de la versión reciente del Stanford-Binet. Como resultado de sus

investigaciones rechazó el concepto de inteligencia general de Spearman y, en cambio, propuso un modelo de siete habilidades mentales primarias: comprensión verbal (v), fluidez verbal (n), aptitud numérica (n), aptitud espacial (s), memoria asociativa (m), velocidad perceptual (p) y razonamiento/inducción (r/i); diferentes individuos poseen cada habilidad en diferentes grados, lo que determina su CI. Propuso que incluso con variables no cuantitativas se puede realizar medición aplicando el coeficiente de contingencia; a este enfoque le llamó la ley del juicio comparativo, un modelo de medición en el que el sujeto experimental compara dos estímulos y emite un juicio sobre su magnitud, lo que resulta fundamental en psicofísica y en psicometría (Thurstone, 1927; 1938). Su ley del juicio comparativo constituyó una de las bases conceptuales para desarrollar la TRI.

El Mental Measurements Yearbook es un sistema de referencia en psicometría que realiza revisiones de los tests publicados. La primera edición como libro estuvo a cargo de Oscar Krisen Buros y se publicó en 1938 por la Rutgers University Press en Nueva Jersey, Estados Unidos (Buros, 1938). A pesar de lo que su nombre indica, las nuevas ediciones se publican cada tres años. La última edición, de 2021, estuvo a cargo del Buros Center for Testing, a través de la University of Nebraska Press.

John Clemans Flanagan (1906-1996), psicólogo estadounidense, en 1941 fue comisionado por el ejército para desarrollar pruebas que se aplicaban para seleccionar pilotos de combate. Desarrolló la técnica del incidente crítico y a partir de ella construyó diversos tests que incluyen habilidades como aritmética, montaje de componentes, memoria, patrones y otras; diseñadas a partir de comportamientos críticos para el desempeño de ocupaciones específicas, permiten clasificar a los sujetos conforme a su



Un nouvel esthésiomètre [Grabado], por Binet, A., 1900, En IX. Un nouvel esthésiomètre (p. 233). *Année Psychologique* 7(1), 231-239. ([URL](#)). Dominio público.

aptitud. Aunque originalmente se aplicaron a ocupaciones militares, estas pruebas posteriormente se llevaron a la selección de personal en las más diversas áreas, desde la industria hasta la mercadotecnia y la educación (Flanagan, 1954).

Starke Rosecans Hathaway (1903-1984), psicólogo, y John Charnley McKinley (1891-1950), neurólogo, ambos estadounidenses, desde la década de los 1930 desarrollaron el Inventario Multifásico de Personalidad de Minnesota (MMPI), un instrumento que evalúa rasgos para la detección de trastornos de salud mental, con el propósito de orientar la intervención psicoterapéutica (Hathaway & McKinley, 1940). La primera versión (MMPI-1) se publicó en 1943 y desde entonces se han desarrollado varias versiones, que incluyen cientos de reactivos; el MMPI-2 alcanzó los 567 reactivos y la versión 2008 los redujo a 338. Los reactivos son del tipo falso-verdadero y se estructuran en diez escalas clínicas, más tres escalas llamadas de validación



Alfred Binet [Fotografía], 1898, Wellcome Collection ([URL](#)). Dominio público.

(que realmente estiman el nivel de verdad mediante la consistencia de las respuestas emitidas por el sujeto examinado), además, la continua investigación con el instrumento ha desarrollado decenas de escalas derivadas.

Frederic Mather Lord (1912-2000), psicómetra estadounidense que trabajó como investigador en el Educational Testing Service, realizó su tesis bajo la dirección de Gulliksen y Tucker. Como producto, publicó en 1953 Una teoría sobre las puntuaciones de los test, que se reconoce como la primera formalización de la Teoría de la Respuesta al Item (Lord, 1952)

Stanley Smith Stevens (1906-1973), psicólogo estadounidense, fue un continuador de los trabajos de Weber y Fechner y propuso la Ley de Potencia de Stevens (Stevens Power Law) que modifica la Ley de Weber-Fechner y constituye la base misma de la psicofísica, al modelar matemáticamente la relación entre los estímulos físicos y su intensidad percibida (Stevens, 1957). Formuló, además, los cuatro niveles o escalas de medición (nominal, ordinal, intervalar y de razón), que son los más aceptados y aplicados en la psicometría actual (Stevens, 1946).

En 1950, Harold Oliver Gulliksen (1903-1996), psicólogo estadounidense que formalizó la TCT a partir del análisis matemático del puntaje verdadero y el error, así como las propiedades de confiabilidad en función de la longitud del test y de validez mediante la selección univariada o multivariada de los reactivos (Gulliksen, 1950). Con la obra de Gulliksen, el desarrollo de los tests avanza desde los ejercicios principalmente empíricos hacia una teoría psicométrica que fundamenta las propiedades del test y contribuye a la comprensión de las dificultades que implica la medición justa y precisa de las variables psicológicas, en cualquier área de aplicación de la psicología.

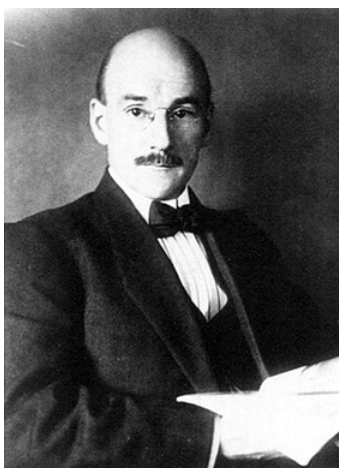


En 1958 Warren Stanley Torgerson (1924-1999), psicólogo estadounidense, desarrolló métodos de medición mediante el escalamiento, que van desde la estimación subjetiva mediante percepciones y opiniones, pasando por el fraccionamiento en partes más pequeñas, hasta el seccionamiento en partes iguales (Torgerson, 1958). Con la contribución de Torgerson se avanza desde las restricciones de la medición clásica en la física hacia la medición de variables psicológicas, que no pueden ajustarse a tales restricciones.

Robert Ladd Thorndike (1910-1990), psicólogo educativo estadounidense, como su padre Edward Lee Thorndike, realizó diversos experimentos sobre aprendizaje con animales. A partir del Lorge-Thorndike Intelligence Test publicado en 1954, desarrolló el Test de Habilidades Cognitivas, un instrumento estandarizado que mide razonamiento y solución de problemas y es aplicable desde preescolar hasta preuniversidad; ha tenido diversas revisiones y se han desarrollado varias formas; aun se aplica ampliamente en la actualidad. Fue un precursor del análisis de grupos (cluster analysis), una técnica estadística que permite agrupar objetos conforme a determinados criterios de semejanza, que hoy se aplica desde el reconocimiento de patrones hasta el aprendizaje automático (machine-learning) y la inteligencia artificial (Thorndike, 1950).

Desde principios de la década de los 1960, en Europa y América creció un movimiento crítico a la TCT por considerar que su dependencia del grupo de referencia en todo el proceso, desde la construcción del test hasta la interpretación de los puntajes, es una fuerte limitante. George

William Rasch (1901-1980), matemático danés que trabajó en psicometría, en 1960 publicó Probabilistic models for some intelligence and attainment tests en el Instituto Pedagógico de Dinamarca, obra que constituyó el primer volumen de la serie Psicología matemática. En la obra de Rasch, que posteriormente fue revisada y aumentada, se desarrolla



Henry, H. Goddard
[Fotografía], Wikimedia Commons ([URL](#)). Dominio público.

formalmente la primera -y más influyente- versión de la TRI como alternativa a la TCT (Rasch, 1980). La TRI se fundamenta en modelos probabilísticos, asume respuestas binarias del tipo falso-verdadero y desplaza el centro de interés de la población al individuo y, consecuentemente, del test al reactivo (ítem). Inicialmente se aplicó a la evaluación de la inteligencia y del aprendizaje, pero una intensa investigación la ha llevado al estudio de prácticamente todos los constructos psicológicos (que la TRI llama “rasgos latentes”). Conforme ha evolucionado, a la TRI (en inglés

ITR, Item Response Theory) en español se le ha llamado Teoría de la Respuesta al Ítem, Teoría de la Respuesta al Reactivo, Teoría de la Curva Característica del Ítem o Teoría del Rasgo Latente.

Robert Glaser (1921-2012), psicólogo estadounidense e investigador en el área educativa, a principios de los 1960 observó que se debe distinguir entre la medición del aprendizaje obtenida a través de los tests y, lo que es un proceso posterior, la evaluación que consiste en la interpretación de la medida con diversos fines (acciones de remedio, promoción, etc.). Propuso una distinción que se volvería clásica entre la evaluación por criterio, que decide sobre el grado de competencia del estudiante con el fin principal de promoverlo a



Sobre todo en el siglo que corre, no puede comprenderse el estado del arte en el campo de la psicometría sin tener al menos una panorámica del desarrollo de la computación. En 1936 se formula la Máquina de Turing, el primer modelo teórico de una computadora (Turing, 1937). En 1943 se publica el primer artículo sobre redes neuronales (McCulloch & Pitts, 1943). En 1946 se presenta ENIAC, la primera computadora electrónica (Goldstine & Goldstine, 1946). A partir de ahí inicia un desarrollo tecnológico que no ha cesado de avanzar y transformar todos los campos de la actividad humana. En 1956, en el Dartmouth College, se realiza el primer proyecto de investigación sobre inteligencia artificial (McCarthy, Minsky, Rochester, & Shannon, 2006). En 1958 Frank Rosenblatt (1928-1971) crea el perceptrón, la primera neurona artificial operativa capaz de

Siglo XXI

seleccionar subgrupos de un grupo de objetos (Rosenblatt, 1958). En 1966, Joseph Weizenbaum (1923-2008) en el Massachusetts Institute of Technology crea Eliza, el primer chatbot con inteligencia artificial que simula un proceso psicoterapéutico (Weizenbaum, 1966). A principios de los 1970, en la Stanford University se crea MYCIN, el primer sistema experto para el diagnóstico de enfermedades infecciosas de la sangre (Shortliffe, 1976). A partir de 1981 se comienza a difundir el uso de las bases de datos inteligentes, la robótica, las redes neuronales, el reconocimiento de patrones, el procesamiento del lenguaje natural; todo ello llevó a la aplicación de la IA. A finales de 2018 OpenAI sacó al mercado las primeras versiones de ChatGPT, un robot virtual capaz de sostener una conversación con un ser humano, generando texto coherente y contextual; en pocos meses aparecieron otros chatbot (robots conversacionales). Y este es el estado del arte en computación a principios de la tercera década del siglo XXI.

En el ámbito de los CAT, en 2010 se funda la Asociación Internacional para los Tests Computarizados Adaptativos, que reúne a los principales expertos en el área y edita el Journal of Computerized Adaptive Testing. En 2020, la APA aprobó los Guidelines for Psychological Assessment and Evaluation que, dentro de sus lineamientos para la práctica psicométrica, ya incluyen algunas consideraciones al respecto del uso de la computación y las comunicaciones en la psicometría. Estos lineamientos tienen como antecedente el documento Technical recommendations for psychological tests and



Lewis Terman [Fotografía], en Store norske leksikon ([URL](#)). Dominio público.



diagnostic techniques, publicado por la APA en 1954.

El interés por métodos más eficientes y precisos, que reduzcan los sesgos en la aplicación de pruebas psicométricas, no es algo nuevo. Ya en 1926 Sidney Leavitt Pressey presentó el prototipo de una máquina para automatizar la presentación de reactivos para medir la inteligencia y estaba interesado en aplicarlas a la automatización de la enseñanza. Años más tarde, Skinner (1958) retomó esta propuesta para desarrollar sus máquinas de enseñanza.

Hubo que esperar varias décadas para que con la popularización de la computadora personal fuera posible implementar los CAT a mayor escala. El trabajo pionero que demostró este avance fue el el Armed Services Vocational Aptitude Battery (ASVAB), el test de aptitud para el ingreso al ejército. Dicha prueba existe en lápiz y papel desde 1968, y en 1979 el Departamento de Defensa comenzó a desarrollar el proyecto de investigación para estandarizarlo en su versión CAT. Su primera versión se puso en funcionamiento a mediados de 1980 (van Der Linden, 2000; Sands, Waters, McBride, 1997). Sin embargo, por el alto costo que implicaba y las limitaciones en la capacidad de cómputo, fue hasta 1996 que se aplicó por primera vez a gran escala (ASVAB, 2023).

El National Council Licensure Examination (NCLEX), es un test de habilitación profesional para enfermeros, se aplica en Estados Unidos, Canadá y Australia. A partir de 1994 se implementó su versión NCLEX/CAT (National Council of State Boards of Nursing, 2014).

El GRE, desarrollado por el Educational Testing Service, desde 1949 se aplica como un test de aptitud para ingresar al posgrado. En 1993 comenzó su aplicación en versión CAT (van der Linden, 2000).

A mediados de 2022 se dio a conocer la iniciativa del ETS por aplicar el SAT en versión CAT. La primera aplicación del CAT-SAT fue a inicios de 2023. Actualmente se distingue entre el Digital SAT (versión adaptativa) y el Full-length Linear SAT (versión no adaptativa) (College Board, 2022).

En este contexto, podemos identificar tres líneas de desarrollo tecnológico en el campo de la psicometría con aplicaciones directas de la computadora.

1. Tests automatizados. El paso del siglo XX al XXI presencié el gradual abandono de los tests del tipo lápiz y papel, para pasar al paradigma digital. Particularmente en las aplicaciones masivas, como es el caso de los exámenes de admisión a la universidad, cada vez más se utilizan tests cuyos reactivos se presentan y responden frente a una computadora, posteriormente se califican y los resultados se emiten también en forma computarizada (Bennett, 1998). Las ventajas son múltiples y evidentes, pero destacan su precisión al calificar, la economía, rapidez en la entrega de resultados y, no menor, la sostenibilidad ecológica. Por otro lado, la seguridad es el problema principal de los tests automatizados; en la mayoría de los casos es necesario aplicarlos en lugares controlados (Test Center) para evitar el fraude (Rojas-Tejada, 2001).

2. Tests adaptativos informatizados (CAT, Computerized Adaptive Testing). Como ya vimos, la TRI no es nueva; sin embargo, solo el gran poder de cómputo disponible actualmente ha permitido desarrollar tests que se adaptan a las respuestas que va dando el examinado a cada reactivo durante la aplicación. A partir de grandes bancos de reactivos que buscan medir cada elemento del constructo que se está evaluando, los CAT son capaces de seleccionar reactivos conforme a las respuestas que cada



sujeto examinado va emitiendo en los reactivos anteriores; en este sentido son adaptativos (van der Linden & Glas, 2000). Los CAT pueden aplicarse a la medición de cualquier constructo psicológico (aprendizaje, personalidad, inteligencia, etc.) y a diversos propósitos (promoción, selección, diagnóstico clínico, encuestas, etc.). Es claro que los tests

adaptativos exigen una enorme capacidad de almacenamiento y procesamiento para tratar con grandes bancos de reactivos; por tanto, solo la computadora los hizo viables.

De este modo, el concepto mismo de test se ha transformado radicalmente, desde un instrumento preconstruido, que tiene una



Group examination in a hospital ward, Camp Lee, October 1917 [Fotografía], Yerkes, R., Flickr ([URL](#)). Dominio público.



selección y una estructura de reactivos fijas, hasta un proceso dinámico en que el sistema va seleccionando y presentando reactivos conforme a las respuestas que va dando cada sujeto individual. Es evidente que las numerosas decisiones y acciones que implica un CAT, exigen una enorme capacidad de almacenamiento y velocidad de procesamiento para tratar con grandes bancos de reactivos y tomar decisiones mediante complejos algoritmos. Nuevamente, todo esto no sería posible sin la computadora.

3. Inteligencia artificial. Esta es la frontera del conocimiento en psicometría. La sinergia entre computación y comunicaciones, así como el acceso a grandes bancos de información (se habla, incluso, de bancos de conocimiento), combinado con el amplio arsenal de instrumentos psicométricos disponibles para medir las más diversas variables en condiciones igualmente diversas, ha modificado radicalmente el paradigma psicométrico para pasar desde un solo test clásico hasta una batería de tests adaptativos que pueden ser aplicados como parte de una intrincada ruta de evaluación a nivel de un solo individuo y, para aumentar aun más la complejidad, en todo esto el psicólogo puede ser auxiliado por un sistema experto, capaz de manejar bases de datos enormes.

Cualquier proceso de evaluación psicológica tiene como propósito final tomar decisiones fundamentadas, que permitan la posterior intervención. En los diferentes campos de la psicología, trátase de la selección de estudiantes o de personal, el diagnóstico clínico o cualquier otro, primero es necesario seleccionar instrumentos adecuados a cada caso particular; se sabe que mientras más amplia sea la batería

psicométrica, mayor serán la confiabilidad y validez de la medición; después hay que aplicarlos, calificarlos e interpretar los resultados. Todo esto puede resultar en un costo prohibitivo. Es aquí donde la inteligencia artificial promete grandes ventajas para los estudios psicométricos.

Sin embargo, debe estar claro que toda aplicación de la inteligencia artificial no puede tener más que un carácter instrumental, es siempre el experto humano quien debe tomar la decisión y asumir la responsabilidad. Para comprender la trascendencia de este principio axiológico, conviene observar un hecho ocurrido hace pocas décadas. El sistema experto MYCIN, arriba citado, en pruebas experimentales demostró un nivel de acierto superior al promedio de los médicos humanos; sin embargo, por cuestiones éticas nunca se autorizó su aplicación, ni siquiera en zonas con carencias de servicios médicos. El principal elemento de juicio que llevó a esta decisión fue el cuestionamiento: en caso de un error de diagnóstico que condujera a provocar daños en el paciente, ¿quién sería responsable, el médico o el sistema?

Terminamos este bosquejo histórico advirtiendo que estamos demasiado cerca de los acontecimientos para poder prever el rumbo que tomarán los cambios en la práctica psicométrica y, sobre todo, los efectos que estos tendrán sobre la práctica de la psicología. Finalmente, como decía Groucho Marx: la predicción es muy difícil, sobre todo si se trata del futuro.



Por lo que se sabe, al menos desde los inicios de las grandes organizaciones sociales fue necesario evaluar las diferencias individuales en aptitud y competencia para seleccionar a los mejores candidatos a ocupar puestos importantes para la administración y la defensa. Tal fue el caso de China, ya sea que haya ocurrido hace dos milenios o hace cuatro, el hecho es que uno de los imperios más longevos, y seguramente el más estable de la historia, para su funcionamiento eficaz dependió en buena medida de la selección de la persona adecuada para ocupar el puesto adecuado en el momento adecuado, principio básico de la selección de personal. Desde los primeros exámenes imperiales, que llegaron a ser supervisados por el emperador en persona en la Ciudad Prohibida de Pekín, hasta la última aplicación del Gaokao, examen de ingreso a la universidad que en 2023 involucró a más de doce millones de aspirantes en una sola fecha, China ha practicado a lo largo de milenios lo que hoy llamamos psicometría.

En Occidente, ya en la Biblia se habla de la selección de personal mediante una prueba (que hoy llamaríamos proyectiva) de conducta abierta. Los filósofos de la Grecia clásica propusieron observar sistemáticamente la conducta abierta para inferir las actitudes y las aptitudes como medio para evaluar la calidad ciudadana; en cuanto al servicio civil de carrera y, sobre todo, al servicio militar, desarrollaron una serie de pruebas de ejecución para evaluar variables físicas e intelectuales consideradas importantes en las personas. Roma, una sociedad que descansó en el ejército durante toda su historia, naturalmente aplicó la evaluación a cuestiones militares mediante la observación, la ejecución y el desempeño.

En la Edad Media, la desintegración del territorio imperial de Roma en pequeños reinos y feudos llevó a una suspensión de la evaluación

Conclusiones

sistemática de las diferencias individuales; es hasta el segundo periodo, la Baja Edad Media, con el surgimiento de las primeras universidades, que la evaluación de variables, ahora de naturaleza principalmente cognitiva, comenzó a desarrollarse y se formaron entonces muchas de las tradiciones que actualmente prevalecen en las universidades; de este modo, el área de aplicación más importante de la psicometría pasó a ser la educación.

En la Edad Moderna surgieron las nuevas ciencias y se aplicaron al desarrollo tecnológico. En el campo de la educación se comenzó a desarrollar el diseño de sistemas de enseñanza eficaces y de aplicación masiva, en los cuales el proceso de evaluación del aprendizaje jugó un papel central. Desde finales del siglo XVIII, los astrónomos, que contaban con telescopios cada vez más potentes, encontraron que las limitaciones del observador, de naturaleza necesariamente individual, jugaban un papel cada vez más importante en la precisión de las observaciones; como respuesta, fue formulada la ecuación personal, primera expresión matemática que toma como objeto central a las limitaciones perceptuales del individuo. A partir de aquí creció el interés por el estudio de los tiempos de reacción, interesados en medir la velocidad del arco sensorio-motor sin tratar de explicar lo que ocurre en el procesamiento cerebral; en esta dirección, destacan los trabajos de Donders que involucraron procesos cognitivos de creciente complejidad, aunque sin tratar de explicarlos.

Ya en el siglo XIX, los experimentos pioneros de Weber y su continuación y formulación



matemática por Fechner, condujeron al nacimiento de la psicología científica y a la creación del primer laboratorio experimental de psicología. En esta época destaca la contribución de Ebbinghaus, que por primera vez logró medir la memoria, utilizando simplemente sílabas sin sentido; todo ello ocurrió en Leipzig. En una línea diferente, en Inglaterra, Galton impulsó la biometría, aplicó la estadística y comenzó a desarrollar métodos matemáticos aplicables a las ciencias de la conducta, poniendo las bases para la construcción de los primeros tests.

Y así se llegó al siglo XX, sin duda la época de mayor desarrollo de la psicometría. De los experimentos y mediciones psicofísicas que privilegiaron la precisión, se pasó a los esfuerzos por medir constructos psicológicos; es decir, se trató de buscar dentro de la caja negra. Fue en Francia, en 1905, donde Binet y Simon construyen el primer test -como hoy se le entiende comúnmente- para evaluar el desarrollo mental de niños escolares; el Test de Binet-Simon tuvo tal éxito que pocos años después fue llevado a Norteamérica, traducido, aplicado masivamente y después adaptado; el ahora llamado Stanford-Binet es un test que aun hoy se aplica. En Estados Unidos, durante el siglo XX creció el interés por aplicar la psicometría a la educación, desarrollando multitud de test que hoy son clásicos y de aplicación masiva, baste recordar como ejemplos al SAT y al GRE. Pero conviene no pasar por alto el desarrollo de los laboratorios de psicología; pocos años después de Leipzig se crearon los primeros laboratorios en Estados Unidos y a partir de ahí, con el desarrollo tecnológico, el centro de gravedad de la psicología pasó a esta región; los laboratorios han jugado un papel central en el desarrollo de la psicometría.

Así, los instrumentos, comúnmente llamados tests, que miden constructos (rasgos latentes, se

les llamaría más tarde) tales como la inteligencia, memoria o ansiedad, desplazaron a la psicofísica y los tiempos de reacción como objetos de interés, a tal grado que ha llegado a entenderse que los tests son el objeto único de la psicometría. Pero esto deberá discutirse en otro lugar.

En el siglo XX, los dos campos en los que más se desarrolló la psicometría fueron la educación y el ejército. Respecto al primero, ya se dijo que durante todo el siglo se desarrollaron tests para medir aptitudes y competencias en el campo de la educación. Respecto al ejército, las dos Guerras Mundiales dieron un impulso definitivo a la psicometría (como a casi cualquier otro campo de la ciencia), a tal grado que teorías e instrumentos clave se desarrollaron como parte de los esfuerzos de guerra.

Después de la psicofísica y la biometría, son dos las teorías fundamentales en la psicometría: la Teoría Clásica del Test (TCT) y la Teoría de la Respuesta al Ítem (TRI). Desde luego, este no es el lugar para intentar una discusión de estas teorías, baste decir que ambas son tan diferentes que, sencillamente, producen instrumentos radicalmente distintos.

Ya en el siglo XXI, la psicometría ha sido tan influida por la computación que simplemente no se entiende a la primera sin la segunda. Los tests computarizados, los tests adaptativos y las potenciales (aunque cada vez más actuales) aplicaciones de la inteligencia artificial, abren caminos para la psicometría, y en consecuencia para la psicología, que la han transformado radicalmente.

*

Estamos en un proceso de cambio tan acelerado, que todo es provisional. Sin embargo, nos atreveremos a identificar cuatro grandes retos que la psicometría habrá de afrontar en los próximos años.

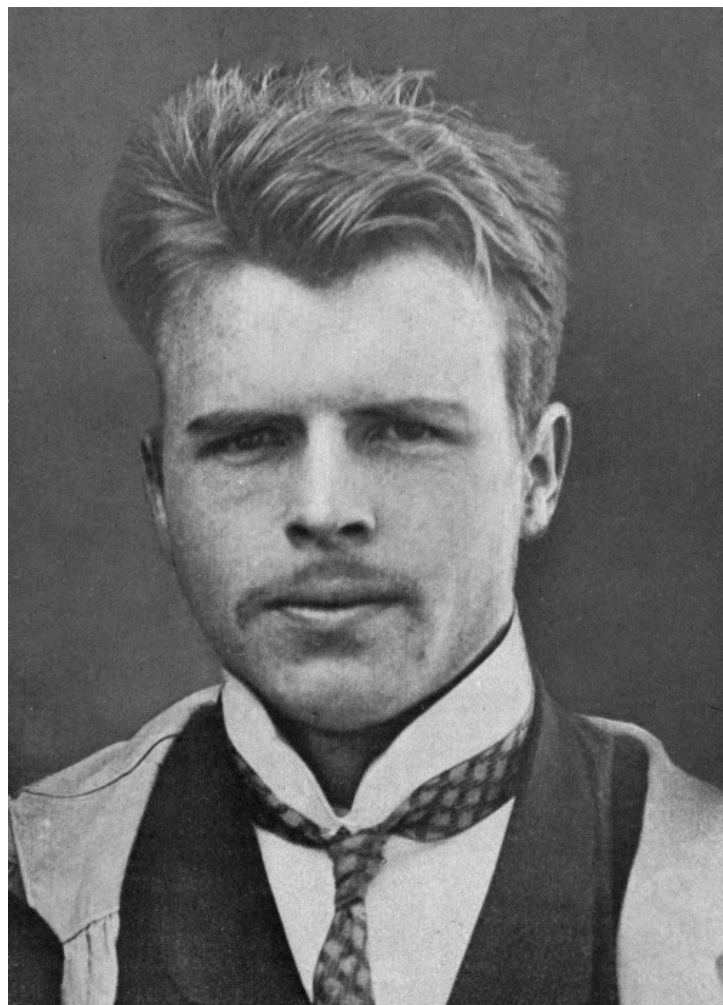


Las aplicaciones tradicionales, individuales o en grupo, basadas en el lápiz y el papel, tendrán que dejar su lugar a aplicaciones computarizadas, cada vez más flexibles y masivas. En consecuencia, el primer reto es la seguridad del banco de reactivos. Si durante el siglo anterior una prioridad fue el control físico de los tests para evitar que el examinado los conociera antes de la aplicación formal, lo que comprometería su validez, en el mundo digital es probable que se pase del test al reactivo. En consecuencia, la seguridad de los bancos de reactivos tendrá en la psicometría la misma importancia que la seguridad de la información bancaria tiene para el sistema financiero. Cómo lograr la seguridad en aplicaciones a distancia, es realmente un reto.

Segundo, la precisión, entendida como la ausencia relativa de errores a lo largo de todo el proceso, puede incrementarse notablemente aprovechando el poder de cálculo de la computadora y dejando al humano la creatividad y la logística para el control de errores en todo el proceso psicométrico. Sin embargo, ya se observó que el uso de la tecnología computacional, particularmente la inteligencia artificial, plantea serios retos de tipo ético, profesional e incluso legal.

Tercero, seleccionar y controlar los contenidos que se presentarán al examinado es una parte central de la validez del test. Puesto que el universo de contenidos puede ser enorme y la necesidad de formas equivalentes lo aumenta aun más, la computadora y en especial la inteligencia artificial pueden convertirse en herramientas que potencien las capacidades humanas para esta gigantesca tarea de análisis. Este enorme poder de medición exigirá del psicólogo, quien no olvidemos es el responsable final de todo el proceso, nuevas competencias, desconocidas en el siglo anterior.

El cuarto y último reto consiste en el mantenimiento y revisión continua que requiere cualquier instrumento psicométrico para evitar la obsolescencia. Si revisamos la forma en que se construían, aplicaban y calificaban los tests en la primera mitad del siglo XX, veremos que la psicometría actual se parece a su antecesora como los ferrocarriles de vapor se parecen a los de levitación magnética. Pero conforme crece la complejidad de un sistema, crece la probabilidad de fallos que pueden llegar a ser catastróficos; por ello, el mantenimiento continuo de los sistemas psicométricos reclamará más atención y recursos que los instrumentos anteriores, de la misma forma que ocurre en otras aplicaciones de la computación.



Hermann Rorschach [Fotografía], Wikimedia Commons (URL). Dominio público.



Cronología

- **2200** Probablemente, durante la dinastía Xia, el emperador chino examinaba a sus oficiales cada tres años para promoverlos o destituirlos.

- **150** Surgen los exámenes en China como un concurso de oposición para funcionarios del imperio.

622 Se consolida un sistema de exámenes en China, durante la Dinastía Tang.

1219 Comienza la aplicación de exámenes de leyes en la Universidad de Bolonia.

1575 Huarte publica Examen de ingenios para las ciencias.

1599 Se promulga la Ratio Studiorum, el primer curriculum completo y documentado.

1795 Maskelyne registra por primera vez un caso de error perceptual a partir de los cometidos por su ayudante Kinnebrook.

1823 A partir de las experiencias de Maskelyne, Bessel formula la ecuación personal.

1850 Helmholtz determina la velocidad del impulso nervioso.

1860 Fechner formaliza la ley que rige la relación entre la intensidad del estímulo y los cambios en la sensación.

1864 Hirsch mide tiempos de reacción empleando el cronoscopio.

1865 Donders publica los primeros resultados de las mediciones de tiempos de reacción.

1866 Donders presenta el noematacógrafo y el noematacómetro.

1879 Wundt establece el primer laboratorio de psicología en Leipzig, Alemania.

1883 Stanley Hall organizó el segundo laboratorio de psicología en la Universidad Johns Hopkins.

1884 Apertura del Laboratorio Antropométrico de Galton en Londres.

1885 Ebbinghaus publica Sobre la memoria, donde reporta sus estudios experimentales.

1886 Cattell publica sus estudios sobre tiempos de reacción.

1887 Cattell funda un laboratorio de psicología en la Universidad de Pennsylvania.

1892 Stanley Hall funda la American Psychological Association.

1896 En la Universidad de Columbia, Cattell aplicó un test a 100 estudiantes, interpretando sus puntajes por su posición relativa.

1900 Se funda el College Entrance Examination Board.

1901 El College Board aplica el primer examen a 973 estudiantes en 67 localidades de Estados Unidos y dos de Europa.

1905 Es abolido el sistema de exámenes imperiales en China.

1905 El test Binet-Simon se aplica para medir inteligencia en niños, introduciendo el concepto de edad mental.

1908 Goddard traduce el test de Binet-Simon y lo aplica en Estados Unidos.



1916 Stern formula el concepto de cociente intelectual.

1917 Terman realizó la revisión conocida como Stanford-Binet.

1917 Yerkes desarrolla el test Army-Alpha.

1917 Woodworth desarrolla la Hoja de datos personales, el primer instrumento para medir personalidad.

1921 Terman inicia un programa para detectar, atender y dar seguimiento a los dotados.

1921 Rorschach publica el instrumento que lleva su nombre, siendo el primer test proyectivo publicado.

1926 Se aplica por primera vez el Scholastic Aptitud Test.

1926 Pressey presenta el prototipo de una máquina para automatizar la enseñanza y la evaluación de la inteligencia.

1927 Spearman formula la Teoría bifactorial de la inteligencia.

1935 Se funda la Sociedad Psicométrica en Michigan, Estados Unidos.

1936 Se publica el primer número de la revista Psychometrika.

1938 Thurstone propone un modelo de Habilidades mentales primarias.

Se publica la primera edición del Mental Measurements Yearbook.

1941 Flanagan desarrolló la Técnica del incidente crítico.

1943 Hathaway y McKinley publican la primera versión del MMPI.

1946 Stevens formula los cuatro niveles de medición.

Se crea la ENIAC, la primera computadora electrónica.

1949 Se crea el GRE un examen para el ingreso al posgrado.

1950 Gulliksen formaliza la Teoría Clásica del Test.

1952 Lord formaliza la Teoría de la Respuesta al Ítem en Una teoría sobre las puntuaciones de los test.

Se aplica por primera vez el Gaokao, examen de ingreso a la universidad en China.

1953 Murray y Morgan publican el Test de Apercepción Temática.

1954 Se crea el GMAT, un examen para el posgrado en Administración.

Thurstone desarrolla el Test de Habilidades Cognitivas.

1956 Se realiza el primer proyecto de investigación sobre Inteligencia artificial.

1958 Torgerson desarrolla métodos de medición mediante el escalamiento.

1960 Rasch desarrolla el Modelo logístico de un parámetro, el primer modelo de la TRI.

1963 Glaser distingue entre evaluación por criterio y evaluación por norma.

1964 Se crea el TOEFL para evaluar el dominio de inglés en hablantes no nativos.

1966 Weizenbaum programa a ELIZA, primer sistema inteligente con propósitos psicoterapéuticos



1983 Se comienza a aplicar el ARE test de habilitación profesional para arquitectos.

1996 Se aplica el CAT-ASVAB, el primer test computarizado adaptativo estandarizado.

2010 Se funda la International Association for Computerized Adaptive Testing (IACAT).

2018 OpenAI sacó al mercado las primeras versiones de ChatGPT.

2020 Se publica la última versión de los APA Guidelines for Psychological Assessment and Evaluation.

2023 El SAT es estandarizado para aplicarse en su versión computarizada adaptativa.

Bibliografía

- American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51(2, Pt. 2), 1-38. Obtenido de URL
- American Psychological Association. (2020). *APA Guidelines for Psychological Assessment and Evaluation*. Recuperado el 1 de Diciembre de 2023. Obtenido de URL
- Anastasi, A. (1973). *Tests psicológicos*. Madrid: Aguilar. Obtenido de URL
- Bennett, R. E. (1998). *Reinventing Assessment: speculations on the Future of Large-Scale Educational Testing. A Policy Information Perspective*. Policy Information Center, Educational Testing Services, 1-27. Obtenido de URL
- Bessel, F. W. (1823). *Astronomische Beobachtungen auf der Königlichen Universitäts-Sternwarte in Königsberg: Band 8 [Observaciones astronómicas en el Observatorio Astronómico de la Universidad de Königsberg]* (Vol. 8). Königsberg: Universitäts - Buchhandlung. Obtenido de URL
- Bessel, F. W. (1875). *Persönliche Gleichug bei Durchgangsbeobachtungen [Ecuación personal en observaciones de tránsito]*. En R. Engelmann, *Abhandlungen [Tratados]* (Vol. III, págs. 300-303). Leipzig: Verlag Von Wilhelm Engelmann. Obtenido de URL
- Binet, A. (1904). *A propos de la mesure de l'intelligence [Sobre la medida de la inteligencia]*. *L'année psychologique*, 11, 69-82. Obtenido de URL
- Bowman, M. L. (1989). *Testing Individual Differences in China*. *American Pshycologist*, 44(3), 576-578. Obtenido de URL
- Buros, O. K. (1938). *The nineteen thirty-eight mental measurements yearbook*. New Brunswick, NJ: Rutgers University Press. Obtenido de URL



- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 11(41), 63-65. Obtenido de URL
- Cattell, J. M. (1928). Early Psychological Laboratories. *Science*, 543-548. Obtenido de URL
- URL
- Cattell, J. M., & Livingstone, F. (1896). Physical and mental measurements of the students of Columbia University. *Psychological Review*, 3(6), 618-648. Obtenido de URL
- College Board. (2022). *Transitioning to a Digital SAT: Ensuring Tests and Scores Are Valid, Reliable, and Fair*. Obtenido de URL
- Comenio, J. A. (1632/1998). *Didáctica Magna* (8 ed.). México: Porrúa. Obtenido de URL
- Compañía de Jesús. (1599/2005). *The Ratio Studiorum: The Official Plan for Jesuit Education*. (C. Pavur, Trad.) St. Louis: The Institute of Jesuit Sources. Obtenido de URL
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412-431. Obtenido de URL
- Doyle, K. O. (1974). Theory and practice of ability testing in ancient Greece. *Journal of the History of the Behavioral Sciences*, 10(2), 202-212. Obtenido de URL
- DuBois, P. H. (1964). A test-dominated society: China, 1115 B.C. - 1905 A.D. En *Proceedings of the 1964 Invitational Conference on Testing Problems* (págs. 12-20). Princeton: Educational Testing Services. Obtenido de URL
- Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie* [Sobre la memoria: Investigaciones en psicología experimental]. Leipzig: Verlag Von Duncker & Humblot. Obtenido de URL
- Fechner, G. T. (1860). *Elemente der Psychophysik* [Elementos de Psicofísica]. Leipzig: Druck und Verlag von Breitkopf und Hartel. Obtenido de URL
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358. Obtenido de URL
- Fullerton, G. S., & Cattell, J. M. (1892). *On the perception of small differences, with special reference to the extent, force, and time of movement*. University of Pennsylvania Press. Obtenido de URL
- Galilei, G. (1632). *Massimi sistemi del mondo tolemaico, e copernicano*. Florencia: Landini, Gio. Batista. Obtenido de URL
- Galton, F. (1875). *English Men of Science: Their Nature and Nurture*. New York: D. Appleton and Company. Obtenido de URL
- Galton, F. (1883). *Inquiries Into Human Faculty and Its Development*. London: Macmillan. Obtenido de URL
- Galton, F. (1884). *Anthropometric Laboratory*. London: William Clowes and Sons, Limited. Obtenido de URL
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519-521. Obtenido de URL
- Goddard, H. H. (1908). The Binet and Simon Tests of Intellectual Capacity. *The Training School*, 5(10), 3-9. Obtenido de URL
- Goddard, H. H. (1913). Standard Method for Giving the Binet Test. *The Training School*, 10(2), 23-30. Obtenido de URL
- Goldstine, H. H., & Goldstine, A. (1946). The Electronic Numerical Integrator and Computer (ENIAC). *Mathematical Tables and Other Aids to Computation*, 2(15), 97-110. Obtenido de URL
- Gulliksen, H. (1950). *Theory of Mental Tests*. New York: John Wiley & Sons. Obtenido de URL
- Hathaway, S. R., & McKinley, J. C. (1940). A Multiphasic Personality Schedule (Minnesota): I. Construction of the Schedule. *The Journal of Psychology*, 10(2), 249-254. Obtenido de URL



- Helmholtz, H. (1903). Vorträge und reden [Conferencias y discursos]. Braunschweig: Friedrich Vieweg und Sohn. Obtenido de URL
- Hirsch, A. (1864). Expériences chronoscopiques sur la vitesse des différentes sensations et de la transmission nerveuse [Experimentos cronoscópicos sobre la velocidad de las diferentes sensaciones y la transmisión nerviosa]. Bulletin de la Société des sciences naturelles de Neuchâtel, 100-114. Obtenido de URL
- Huarte, J. (1575/2003). Examen de ingenios para las ciencias. Alicante: Biblioteca Virtual Miguel de Cervantes, 2000. Obtenido de URL
- Husek, W. J. (1969). Implications of Criterion-Referenced Measurement. Journal of Educational Measurement, 6(1), 1-9. Obtenido de URL
- Jenckes, T. (1867). Review of Civil Service of the United States. Report Presented by Mr. Jenckes, from the Joint Select Committee of the Two Houses of Congress Appointed July 19, 1866. The North American Review, 105(217), 478-495. Obtenido de URL
- Kant, I. (1786/1989). Principios metafísicos de la ciencia de la naturaleza. Madrid: Alianza Editorial. Obtenido de URL
- Le Goff, J. (1996). Los intelectuales en la Edad Media. Barcelona: Gedisa. Obtenido de URL
- López-Valdepérez, M. (2018). El ejército romano en la tardorepública. Universidad de Zaragoza. Obtenido de URL
- Lord, F. M. (1952). A Theory of Test Scores. Psychometric Monograph(7), 1-84. Obtenido de URL
- Maskelyne, N. (1799). Astronomical Observations Made at the Royal Observatory at Greenwich from 1787-1798 (Vol. III). London: The Royal Society. Obtenido de URL
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine, 27(4), 12-14. Obtenido de URL
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115-133. Obtenido de URL
- Morgan, W. G. (1995). Origin and History of the Thematic Apperception Test Images. Journal of Personality Assessment, 65(2), 237-254. Obtenido de URL
- National Council of State Boards of Nursing. (2014). The Evolution of the NCLEX: 20 Years as a Computer Adaptive Exam. Obtenido de URL
- Official ASVAB. (2023). The Armed Services Vocational Aptitude Battery (ASVAB). Obtenido de URL
- Pillsbury, W. B. (1947). Biographical Memoir of James McKeen Cattell 1860-1944 (Vol. XXV). National Academy of Sciences. Obtenido de URL
- Popham, W. J., & Husek, T. R. (1969). Implications of Criterion-Referenced Measurement. Journal of Educational Measurement, 6(1), 1-9. Obtenido de URL, URL
- Pressey, S. L. (1926). A Simple Apparatus which gives Tests and Scores-and Teaches. School and Society, 23(586), 373-376. Obtenido de URL
- Psychometric Society. (2019). History of The Psychometric Society. Obtenido de URL
- Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press. Obtenido de URL
- Rojas-Tejada, A. J. (2001). Pasado, presente y futuro de los Tests Adaptativos Informatizados: entrevista con Isaac I. Bejar. Psicothema, 13(4), 685-690. Obtenido de URL
- Rorschach, H. (1951). Psychodiagnostics: A Diagnostic Test Based on Perception (5th ed.). (B. Kronenberg, Ed., & P. Lemkau, Trad.) New York, USA: Grune & Stratton Inc. Obtenido de URL
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386-408. Obtenido de URL

- Sands, W. A., Waters, B. K., & McBride, J. R. (1997). *Computerized Adaptive Testing: From Inquiry to Operation*. Washington, DC: American Psychological Association. Obtenido de URL
- Shortliffe, E. H. (1976). *Computer-Based Medical Consultations: MYCIN*. New York: Elsevier. Obtenido de URL
- Skinner, B. F. (1958). *Teaching Machines*. *Science*, 128(3330), 969-977. Obtenido de URL
- Soto-Artuñedo, W. (1999). La "Ratio Studiorum": La pedagogía de la Compañía de Jesús. *Proyección: Teología y mundo actual*, 46(195), 259-276. Obtenido de URL
- Spearman, C. (1927). *The abilities of man, their nature and measurement*. London: McMillan & Co., Limited. Obtenido de URL
- Stern, W. L. (1916). *Der Intelligenzquotient als Maß der kindlichen Intelligenz, insbesondere der Unternormalen* [El cociente intelectual como medida de la inteligencia en los niños, con especial referencia a los subnormales]. *Zeitschrift für angewandte Psychologie*, 1-18. Obtenido de URL
- Stevens, S. S. (1946). *On the Theory of Scales of Measurement*. *Science*, 103(2684), 677-680. Obtenido de URL
- Stevens, S. S. (1957). *On the psychophysical law*. *The Psychological Review*, 153-181. Obtenido de URL
- Terman, L. M. (1926). *Genetic studies of genius, Mental And Physical Traits Of A Thousand Gifted Children* (Vol. I). California: Stanford University Press. Obtenido de URL
- Terman, L. M., Lyman, G., Ordahl, G., Ordahl, L. E., Galbreath, N., & Talbert, W. (1917). *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence*. Baltimore: Warwick & York, Inc. Obtenido de URL
- Thorndike, R. L. (1950). *Individual Differences*. *Annual Review of Psychology*, 1(1), 87-104. Obtenido de URL
- Thurstone, L. L. (1927). *A law of comparative judgement*. *The University of Chicago*, 34(4), 273-286. Obtenido de URL
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press. Obtenido de URL
- Torgerson, W. S. (1958). *Theory and methods of scaling*. London: John Wiley & Sons, Inc. Obtenido de URL
- Turing, A. M. (1936). *On computable numbers, with an application to the Entscheidungsproblem*. *Proceedings of the London Mathematical Society*, 2(1), 230-265. Obtenido de URL
- Van der Linden, W. J., & Glas, G. A. (2000). *Computerized Adaptive Testing: Theory and Practice*. Dordrecht: Kluwer Academic Publishers. Obtenido de URL
- Weber, E. H. (1834). *De Pulsu, resorptione, auditu et tactu: Annotationes anatomicae et physiologicae* [Sobre el pulso, la resorción, la audición y el tacto: Anotaciones anatómicas y fisiológicas]. Leipzig: C. F. Koehler. Obtenido de URL
- Weizenbaum, J. (1966). *ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine*. *Association for Computing Machinery*, 9(1), 36-45. Obtenido de URL
- Woodworth, R. S. (1932). *Robert S. Woodworth*. En C. Murchison, *A history of psychology in autobiography* (Vol. II, págs. 359–380). Worcester: Clark University Press. Obtenido de URL
- Wundt, W. M. (1897). *Grundriss der Psychologie* [Fundamentos de psicología]. Leipzig: Verlag Von Wilhelm Engelman. Obtenido de URL
- Yoakum, C. S., & Yerkes, R. M. (1920). *Army Mental Tests*. New York: Henry Holt and Company. Obtenido de URL



LOS CONTENIDOS DE ÉSTE NUMERO PUEDEN
SER REPRODUCIDOS EN TODO O EN PARTE,
ELECTRÓNICAMENTE O EN FÍSICO, SIEMPRE Y
CUANDO SE CITE LA FUENTE Y SEA CON FINES
ACADÉMICOS Y NO COMERCIALES

adolfolopezsuarez@gmail.com