

Cronología (1822-1988)

- 1.822. Bessel publica un trabajo acerca del "incidente Kinneybrook" sucedido en 1796 en el observatorio de Greenwich, dando lugar al desarrollo de la llamada "ecuación personal".
- 1.835 El estadístico y astrónomo belga Adolphe Quetelet muestra la aplicación del cálculo de probabilidades al análisis de la variabilidad de medidas a la investigación de cuestiones sociales, poniendo de manifiesto que algunas características humanas siguen la curva normal. Quetelet ve las variaciones como errores en torno al hombre medio.
- 1.838 El médico francés Esquirol pone de manifiesto que existen diferentes grados de retraso mental y ensaya diferentes procedimientos para la clasificación de los sujetos retrasados en el grado correspondiente. Entre sus valiosas sugerencias queda el juicio sobre el uso del lenguaje del sujeto como criterio para esa clasificación.
- 1.845 Se produce una innovación en la medición del rendimiento académico en las escuelas públicas de Boston: los exámenes orales se sustituyen por exámenes escritos realizados por inspectores. Este paso va en la línea de someter a todos los sujetos a las mismas condiciones y eliminar la subjetividad.
- 1.848. Seguin elabora pruebas de rendimiento sensorial y motor con el propósito de detectar niños deficientes. Seguin fue un médico francés que rechazó la idea de la incurabilidad de los retrasados mentales, estableciendo en 1837 la primera escuela para niños deficientes. En 1.848 emigró a Estados Unidos consiguiendo que sus ideas se divulgaran. Desde el punto de vista de los tests mentales es conocido especialmente por una prueba manipulativa, el llamado test de ajuste de formas de Seguin, en el que el sujeto debe insertar tan rápidamente como pueda piezas de diversos tamaños en su lugar de ajuste.
- 1.850 El 22 de Octubre Fechner concibe su ley logarítmica que relaciona incrementos geométricos en los estímulos con incrementos aritméticos en las sensaciones de los sujetos.
- 1.860. "Elementos de Psicofísica" (Elemente der Psychophysik) de Gustav Theodor Fechner. Boring (1.950) situará en esta obra el origen de la psicología experimental.

- 1.869. "Classification of Men According to their Natural Gifts" (1883: "Inquiries into Human Faculty and its Development") de Sir F. Galton.
- 1.872 En algunos departamentos gubernamentales de Estados Unidos comienzan a utilizarse exámenes para la selección de acceso. Desde 1.883 esto constituirá la práctica habitual.
- 1.879 Fundación del primer laboratorio de psicología experimental por Wundt en Leipzig.
- 1.884 Galton establece el laboratorio antropométrico en la Exposición Internacional. Pagando tres peniques los sujetos eran medidos en una variedad de formas acerca de sus rasgos físicos, agudeza visual y auditiva, energía muscular, tiempo de reacción y otras cosas semejantes. Por ejemplo, pueden citarse 'la barra de Galton' para la discriminación visual de la longitud, el "silbato de Galton" para determinar el sonido más agudo perceptible. Cuando concluyó la exposición el laboratorio fue trasladado al "South Kensington Museum" de Londres donde siguió funcionando y recopilando datos durante seis años. Galton estaba interesado fundamentalmente en la herencia, influido por la teoría evolucionista de su primo Darwin, y defendió la eugenesia. Galton estaba interesado en la inteligencia, y no en las funciones motrices o sensoriales elementales, sin embargo, de acuerdo con sus propias observaciones y con el espíritu del empirismo de Locke, los sujetos más aptos intelectualmente presentarían mayor capacidad también en estas funciones básicas. Para investigar acerca de estas cuestiones Galton comprendió que era necesario efectuar mediciones. Las mediciones llevaron a una necesidad de tratamiento estadístico de los datos, una línea que desarrolló plenamente su discípulo Pearson.
- 1.887 S.E. Chaille publica en una revista de muy limitada circulación, el New Orleans Medical and Surgical Journal, una serie de test para bebés ordenados en función de las edades en la que la mayoría de los bebés pasaban las pruebas propuestas.
- 1.888 Galton utiliza el término "co-relation". "Co-relations and their measurement, chiefly from Antropometric Data."
- 1.888 James McKeen Cattell ejerce como lector en la Universidad de Cambridge, conociendo a Sir Francis Galton. Cattell se había doctorado en Leipzig con un trabajo acerca de las diferencias individuales en los tiempos de reacción, una línea de investigación que no era del total agrado de W. Wundt. El contacto de Cattell con Galton sirvió para afianzar en el primero su interés por el tema de las diferencias individuales y la necesidad de su medición, desplazando el acento hacia los intereses de Galton más centrados en las facultades superiores.
- 1.888 Inauguración de un laboratorio de Tests en la Universidad de Pennsylvania.
- 1.889 Oehrn, un discípulo de Kraepelin, se interesa por la cuestión de la relación entre las funciones superiores de la mente, para su investigación utiliza tests de percepción, y funciones motoras, pero también tests de memoria.

- 1.890. J.M. Cattell crea la expresión test mental ("Mental tests and measurements"). En este artículo James McKeen Cattell describe una serie de tests que se aplicaban anualmente a los estudiantes universitarios con el propósito de llegar a estimar sus capacidades superiores. Eran tests de aplicación individual fundados en los procedimientos que Cattell había aprendido de Galton e incluían una variedad de aspectos sensomotores. Cattell integró en ellos también medidas del tiempo de reacción, traídas de sus estudios en Leipzig. En todo caso todas estas medidas simples tenían la ventaja de ser manejables, claramente definibles y origen de resultados precisos, mientras que otros procedimientos de estimación de las capacidades superiores, como exámenes orales con juicios subjetivos, resultaban vagos e indeterminados. Durante toda la década J. McKeen Cattell y su discípulo Wissler en la Universidad de Columbia, ponen a prueba diferentes tests mentales en su relación con los niveles en la enseñanza encontrando correlaciones próximas a cero (la más alta fue 0'19). Los tests de Cattell no sólo no estaban relacionados claramente con las calificaciones académicas o las estimaciones de los profesores, tampoco mantenían una correspondencia de unos a otros.
- 1.890 Nace el que sería discípulo de Pearson, R.A. Fisher, en East Finchley.
- 1.892 Kraepelin utiliza tests de asociación libre para indagar el deterioro mental y los efectos de la fatiga.
- 1.893 En Chicago se celebra una exposición donde Jastrow, al modo de Galton, examina a sujetos en tests de tipo sensomotriz. Como aportación, Jastrow muestra a los visitantes como sus puntuaciones quedan en comparación con la "norma" en cada una de esas medidas.
- 1.894 Gilbert en la Universidad de Yale realiza un intento semejante al de J.M. Cattell y encuentra también una relación muy baja entre tests mentales y nivel en la educación.
- 1.894 Sommer sugiere que el test de asociación libre puede usarse también para clasificar a los sujetos patológicos en función de su trastorno mental.
- 1.895 Kraepelin, interesado en la clasificación psiquiátrica, elabora un conjunto de tests basados sobre operaciones aritméticas simples destinados a medir aspectos básicos de caracterización de los sujetos, entre ellos la memoria y la resistencia a la distracción.
- 1.895 Binet y Henri publican un artículo que marca un cambio de rumbo en la elaboración de tests. Hacen una crítica abierta de los procedimientos de medición basados en aspectos sensoriales y motores simples. Sugieren que aunque la medida de las funciones complejas sea menos precisa, la existencia de mayores diferencias entre los sujetos en esas funciones permitiría un diagnóstico relevante. Frente a los tests sensoriales y motrices proponen una dilatada gama de tests centrados en medir la memoria, la imaginación, la atención, la comprensión, la sugestibilidad etc. este trabajo constituye un antecedente claro de la línea de los tests de inteligencia que Binet y Simon emprenderían pronto.

- 1.896 Ferrari y su discípulo Guicciardi se ocupan de la elaboración de tests para el examen de sujetos patológicos. Usan tests motores y sensoriales, pero incluyen también una prueba de interpretación de cuadros.
- 1.897 Ebbinghaus aplica a escolares tests de cálculo aritmético, de memoria inmediata y de tipo verbal con tareas como completar una frase truncada. Entre sus resultados encuentra que solo el test de frases truncadas más complejo muestra una clara similitud con el patrón de rendimiento escolar de esos niños. Este es un resultado importante que comienza a dar apoyo empírico a la idea de utilizar pruebas complejas que supongan el uso de facultades superiores.
- 1.901. Teoría de la correlación de K. Pearson.
- 1.901 Karl Pearson funda la revista *Biometrika*. El primer número aparece en el mes de Octubre.
- 1.904. Spearman formula su modelo incluyendo la teoría del error de medida, la fiabilidad y la fórmula de corrección por atenuación. Casi todas las fórmulas básicas de la teoría clásica de los tests están disponibles desde dos de los trabajos de Spearman publicados en este año y desde sus posteriores aportaciones de 1.907, 1.910 y 1.913. La fórmula de corrección por atenuación sería criticada ese mismo año por Pearson; la cuestión, según la revisión que Thouless realizó en 1.939 parece residir más bien en la interpretación y uso que se haga de la misma.
- 1.905. Test de inteligencia para las tareas de selección de niños en las escuelas de París por Binet y Simón. En 1.904 el ministro de instrucción pública había formado una comisión para tratar acerca de los procedimientos de enseñanza de los niños deficientes. Binet llevaba ya algún tiempo trabajando en la cuestión de como medir la inteligencia, sería equivocado pensar que la línea de los tests mentales que inauguró fuera fruto de la primera intención del autor. Por el contrario Binet también trabajó con medidas de tipo antropométrico, motor y sensorial, y llegó a poner a prueba la quiromancia como método de medición de las capacidades superiores. Fue la evidencia reiterada la que forjó la idea de que la medición directa mediante el enfrentamiento del sujeto con tareas complejas, aunque menos precisas que las medidas sensoriales, ofrecía el tipo de datos relevantes para el propósito de la medición. Binet presentó su escala como una versión provisional de trabajo. No se ofreció ningún procedimiento objetivo para construir una puntuación global final. La que fue llamada "la escala de 1.905" estaba formada por 30 problemas o tests, preferentemente relativos a cuestiones de juicio, comprensión o razonamiento, ordenados por su dificultad de modo creciente. Para determinar este orden de dificultad se aplicaron los tests a cincuenta niños normales de tres a once años de edad y también a algunos niños retrasados. Los treinta tests no excluían tareas sensoriales o motoras, pero el mayor peso estaba en tareas de alto componente verbal u orientadas a problemas "complejos". En 1.908 y en 1.911 se producen nuevas ediciones revisadas de la prueba.

- 1.908 Revisión de la escala de 1.905 de Binet y Simon. Se amplia el número de tests, se eliminan algunos que se había mostrado que no resultaban adecuados y se ordenan todos ellos en niveles de edad desde los 3 a los 13 años. En cada nivel de edad se sitúan aquellas pruebas que los sujetos pueden resolver a esa edad. La puntuación de los sujetos en el test se expresa como su edad mental, el nivel que el sujeto era capaz de llegar a resolver.
- 1.910 En el Volumen 3 del *British Journal of Psychology* aparece por primera vez la fórmula para la estimación de la fiabilidad de un test con longitud doble a uno dado y la de su generalización para longitud n veces la de un test dado. Estas formulas habían sido desarrolladas por separado por Spearman, que las presenta en la página 290 y por William Brown, que las presenta en el artículo siguiente en la página 299. Desde entonces se conocen como fórmulas de Spearman-Brown.
- 1.911 Año de la muerte de Binet. Aparece una tercera revisión de su escala, solo con cambios menores y algunas variaciones en la ordenación de los tests. Lo más notable de esta revisión es que la escala se extiende hasta el nivel adulto. Las versión original de la escala de 1.905 y sus dos revisiones de 1.908 y 1.911 son de aplicación individual y de orientación clínica, para el examen cuidadoso de un solo sujeto en cada uso.
- 1.911 Healy-Fernald presentan una serie de tests de ejecución o manipulativos como complementarios a los de contenido verbal como la escala de Binet.
- 1.912 Stern sugiere el uso de un "coeficiente mental".
- 1.912 Revisión Kuhlmann-Binet de la escala de Binet. Su principal aportación es extender el límite inferior de edad a la que es aplicable la escala de Binet hasta los tres meses.
- 1.914. Primera teoría factorial de la inteligencia de Spearman.
- 1.915. "Test de laberintos" de Porteus que pretende evaluar capacidades mentales sin influencias culturales.
- 1.915. Test de habilidad mecánica de Stenquist.
- 1.916 "Escala Stanford-Binet para la medición de la inteligencia", primera adaptación, tipificación y validación americana del test de Binet-Simon, bajo la dirección de L.M. Terman. En este test se aplica por vez primera el cociente de inteligencia CI, como razón entre la edad mental y la cronológica. La revisión implica tantos cambios que constituye realmente un nuevo test: más de una tercera parte de los "tests" son nuevos y otros son reubicados. Se realiza una tipificación con una muestra americana de mil niños y cuatrocientos adultos. Se hace un esfuerzo por precisar las condiciones e instrucciones de examen.
- 1.917 La APA designa un comité presidido por Robert M. Yerkes para contribuir al ejercito con motivo de la entrada de los estados Unidos en la Primera Gran Guerra. Desde hacía un tiempo Arthur S. Otis había preparado un conjunto

de tests de inteligencia que no había publicado todavía. Esos tests constituyeron el material principal sobre el que se elaboraron las pruebas de inteligencia del ejército americano que tuvieron una contribución inmediata en la tarea de selección y clasificación de un millón y medio de reclutas. Aquellos tests fueron denominados el "Army Alpha" y "Army Beta", dedicándose el segundo, de tipo no verbal, a analfabetos y extranjeros que no dominaban el inglés. Sólo después de la guerra estos tests pudieron ser utilizados con fines civiles. La principal innovación consiste en la aplicación colectiva de los tests que se convierten en un instrumento socialmente útil para la medición de masas de individuos. Tomando como base estos tests colectivos, el uso de los tests se popularizó de un modo ilimitado durante la década siguiente. Prescindiendo del estado de desarrollo técnico de los tests y las precauciones que el profesional tomaría, se usaron tests en todos los contextos colectivos por personal no cualificado obteniendo predicciones y juicios no justificados por las medidas usadas.

- 1.917 Escala de ejecución de Pintner-Paterson, primer intento considerable de crear una escala de tests manipulativos con normas generales tipificadas.
- 1.918 "Personal Data Sheet" primer test colectivo de personalidad creado para una medida del ajuste personal, preparado por Woodworth y usado para seleccionar soldados americanos.
- 1.918 "The 17th Yearbook of the National Society for the Study of Education" es dedicado a una revisión de los tests educativos y psicológicos disponibles.
- 1.919 Edición de la escala de evaluación del desarrollo de las primeras edades de Gesell.
- 1.921 Otis y Knollin, en un trabajo titulado "The reliability of the Binet scale and of pedagogical scales", y también Kelley, por su parte, en ese mismo año, ponen de manifiesto la "superioridad" del error de medida al coeficiente de fiabilidad como indicación de la fiabilidad debido a que el error de medida no cambia con la heterogeneidad del grupo.
- 1.921 Kelley publica por primera vez, en un trabajo titulado "The reliability of tests scores" las ecuaciones que presentan el efecto de la homogeneidad o heterogeneidad del grupo sobre el coeficiente de fiabilidad. Estas ecuaciones fueron desarrolladas bajo el supuesto de que el error de medida se mantiene constante con respecto a las variaciones en capacidad del grupo testado. Holzinger en 1.921, Otis en 1.922 y el mismo Kelley en 1.923 y 1.927 contribuyeron de diversos modos al desarrollo de estas ecuaciones. Holzinger desempeñó un papel crítico poniendo en cuestión el supuesto de equiparación de los errores de medida. Por otra parte Kelley pone de manifiesto que si bien el error de medida no cambia con la variabilidad del grupo, sin embargo la unidad en la que el error de medida se expresa si cambia de un test a otro.
- 1.921 Se publica el "Psicodiagnóstico" de Rorschach con su famoso test basado en la presentación de manchas de tinta.

- 1.922. J.M. Cattell funda la "Psychological Corporation" primera entidad dedicada a la publicación masiva de tests.
- 1.922 L.J. O'Rourke es nombrado director del departamento de investigación en la administración norteamericana. Bajo su dirección se introducen en los procedimientos de selección de personal de la administración las técnicas de tests.
- 1.923. Las fórmulas de Spearman-Brown fueron acogidas con una reacción crítica. Este año se publican sendos trabajos de Holzinger y de Crum poniendo en cuestión aquellas. Holzinger llega a la conclusión de que si se incrementa la longitud del tests más de cinco veces la fórmula sobrestima la fiabilidad a alcanzar. Sin embargo, la reacción crítica que representan los trabajos de Crum y Holzinger fue contrarrestada por los trabajos siguientes que mostraron que la fórmula era razonablemente precisa.
- 1.923. Se publica el Stanford Achievement Test, una batería de rendimiento académico que incorpora los avances de los tests mentales al campo educativo.
- 1.925. R.A. Fisher publica su "Theory of Statistical Estimation" estableciendo las bases del método de la máxima verosimilitud y la estimación puntual de parámetros.
- 1.925 Floyd H. Allport y D. A. Hartman publican "Measurement and Motivation of Atypical Opinion in a certain group" (Amer. Pol. Sci. Rev.; Vol 19; pág 735-760). En el trabajo se utilizan histogramas para describir las frecuencias que reciben diversas opiniones supuestamente ordenadas en una serie. Puede considerarse como uno de los primeros intentos de efectuar mediciones de opiniones. De hecho, cuando tres años después Thurstone se ocupa de la medición de las actitudes habrá de reconocer que su trabajo es una extensión y perfeccionamiento del de Allport y Hartman y que fueron los trabajos de Allport los que suscitaron en él el interés por la cuestión de la medición de opiniones.
- 1.926. Goodenough prepara el test de dibujo de la figura humana.
- 1.926 Peterson publica una relación de las primeras escalas americanas traducción y adaptación de la de Binet. En este trabajo se ofrece también una detallada descripción de las escalas originales de Binet-Simon.
- 1.927. Se presenta el inventario de intereses vocacionales de Strong.
- 1.928 T.L. Kelley publica un estudio acerca de las capacidades mentales diferenciables. Se están dando los pasos hacia la construcción de una concepción factorial de las capacidades con escalas de medición distintas para cada capacidad.
- 1.928 Germain y Rodrigo adaptan al español el Test de medida de la inteligencia de Stanford-Binet.

- 1.928. El instituto Panukraniano de la URSS publica trabajos sobre la medida de la inteligencia.
- 1.928. Louis Leon Thurstone fundamenta las escalas de medición de actitudes en un artículo histórico "Attitudes can be measured" publicado en el *American Journal of Sociology* (Vol 33; pag 529-554). El artículo dedica una parte de su contenido a mostrar que es posible medir actitudes y que ello tiene significado y utilidad. Después describe el procedimiento de elaboración de una escala con intervalos aparentemente iguales detallando cada uno de los pasos y estableciendo los principios de ambigüedad e irrelevancia de enunciados.
- 1.930 Se funda el Australian Council for Educational Research bajo una subvención de la Carnegie Corporation de Nueva York. Uno de los temas a los que el ACER dedicará mayor atención será la medición en educación y psicología.
- 1.931. Louis Leon Thurstone "Análisis factorial Múltiple".
- 1.931 Paulsen sugiere métodos para determinar la inestabilidad del rasgo subyacente medido separadamente de la inestabilidad del test que se usa para medirlo. Este trabajo de Paulsen será continuado por los de Thouless en 1.936 y en 1.939.
- 1.932 T.F. Lentz, Hirshstein, Bertha y J.H. Finch publican en el *Journal of Educational Psychology* "Evaluation of methods of evaluating tests items" un trabajo que puede considerarse pionero en la comparación de métodos en el campo del análisis de ítems.
- 1.932 Lincoln pone de manifiesto que la fiabilidad puede ser muy alta aún cuando las diferencias entre dos conjuntos de medidas supuestamente paralelas sean muy grandes. Lincoln volverá a ocuparse de esta cuestión en un nuevo trabajo en el año 1.933, y también Ackerson en ese año volverá sobre el tema.
- 1.932 R.A. Likert publica en "Archives of Psychology" (Vol. 140) el trabajo "A technique for the measurement of attitudes". Este trabajo implica la divulgación de la estrategia de "respuesta-directa" en escalamiento, donde los sujetos evalúan directamente sobre una escala de anclajes de acuerdo o desacuerdo su relación con un ítem. La sencillez y la utilidad directa del procedimiento pueden contribuir al dilatado y extenso uso de esta técnica de medición.
- 1.934. J.L. Moreno publica "Who Shall Survive?" conteniendo una primera aproximación a la sociometría que ha sido considerada como fundacional.
- 1.935 Se funda la "Psychometric Society".
- 1.935. Johnson. Máquina para la aplicación automatizada de tests.
- 1.935 L.L. Thurstone publica "Vectors of Mind: Multiple-factor analysis for the isolation of primary traits"

- 1.935 Long y Sandiford en un trabajo titulado "The validation of tests items" informan de veintitres índices distintos para el análisis de ítems.
- 1.936 Primer volumen de la revista Psychometrika. Presentada bajo el lema "A journal devoted to the Development of Psychology as a Quantitative Rational Science."
- 1.936 Guilford en un trabajo publicado en Psychometrika titulado "The determination of ítem difficulty when chance success is a factor" desarrolla la fórmula de corrección del azar en la respuesta a ítems que actualmente seguimos usando.
- 1.936 Abolición de los tests mentales en la URSS.
- 1.936 Se publica la primera escala de desarrollo social, la "Vineland Social Maturity Scale" por Doll.
- 1.936. Primera edición del "Mental Measurement Yearbook" de Buros. Le seguirán sucesivas ediciones anuales entre las que pueden destacarse la de 1.937, 1.938, 1.941, 1.949 y 1.965 en las que se repasan los tests psicológicos y educativos disponibles. Los anuarios de tests de Buros comprenden casi la totalidad de tests disponibles en el mercado en lengua inglesa. Cada anuario recoge los tests publicados en una época determinada de modo que completa en lugar de eliminar a los anuarios anteriores. Las primeras ediciones son simples bibliografías de tests, pero desde 1.938 el anuario esta concebido de modo que contiene revisiones críticas de los tests y la bibliografía acerca de las publicaciones sobre cada uno de los tests.
- 1.936 Guilford publica "Psychometric Methods" donde presenta una descripción de noventa métodos de análisis de ítems desarrollados en los años anteriores, buena parte de ellos sin suficiente soporte teórico o metodológico. Esta lista, a pesar del incrementado número de índices, no difiere sustancialmente de la presentada un año antes por Long y Sandiford.
- 1.937 Terman y Merrill publican, como fruto de unos diez años de investigaciones, la segunda revisión del test de Stanford. Consta ahora de dos formas equivalentes, la forma L y la forma M. Se extienden las escalas y se vuelve a tipificar con una muestra de tres mil ciento ochenta y cuatro sujetos desde año y medio hasta los dieciocho años, con el mismo número de chicos que de chicas en cada edad, en diferentes comunidades y controlando exactamente los meses que los sujetos pasaban de su aniversario.
- 1.937 G. F. Kuder y M. W. Richardson, de la Universidad de Chicago, publican un artículo en el número 3 del volumen 2 de Psychometrika titulado "The theory of the estimation of test reliability" donde resaltan que el coeficiente de fiabilidad es "solo una estimación del porcentaje de varianza total que puede describirse como varianza verdadera, es decir, no debida a error" , ponen de manifiesto las dificultades asociadas al cálculo del tests retest y al uso del método de las dos mitades, que puede arrojar muy diversos resultados, poniendo la noción de equivalencia en el centro de la discusión. Desde este punto discuten diversos casos en que, a partir de diversas

informaciones, pueda calcularse un único estimador de la fiabilidad del test en función de los datos de una sola aplicación. La fórmula 20 de este artículo quedaría para siempre como la KR20, expresada como

$$r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma_t^2 - npq}{\sigma_t^2}$$

en su notación y forma original, tal como aquí la presentamos, "la ecuación (20) da una estimación de la fiabilidad de un test, conocido el número de items, la desviación estandar y la varianza *media* de los items." (Cursiva en el original).

- 1.938 Publicación del Test de Apercepción Temática (TAT) por Murray y colaboradores.
- 1.938 Publicación de los trabajos de Richardson "Multidimensional Psychophysics" y de Young y Householder "Discussion of a set of points in terms of their mutual distances" que pueden considerarse el comienzo del escalamiento multidimensional (MDS). Sin embargo hasta la publicación de los trabajos de Shepard (1.962) y de Kruskal (1.964) sobre escalamiento multidimensional no métrico no comenzarán las aplicaciones en psicología de esta técnica.
- 1.938. Thurstone "Primary Mental Abilities".
- 1.939 Hildreth publica una bibliografía sobre los tests psicológicos aparecidos en los cincuenta años anteriores conteniendo valiosas referencias acerca de los tests pioneros. En 1.946 Hildreth publicará un suplemento ampliando su bibliografía hasta contener el año 1.945.
- 1.939 Publicación de la primera versión del Weschsler-Bellevue orientada a la medida de la inteligencia de adultos.
- 1.939 Se publican las tablas de Taylor-Russell que pueden considerarse un ejemplo precursor de la aplicación de la teoría de la decisión a los test. Con estas tablas, a partir del coeficiente de validez del test, del índice de selección (proporción de solicitantes a admitir) y del índice basal (de la proporción de solicitantes aptos que se seleccionarían sin el test), puede calcularse la ganancia en la selección que produce el test debido a su nivel de validez para esas condiciones. Diversos autores prosiguieron el estudio de esta cuestión estableciendo en general (p.e Brogden, 1.946) que la mejora es proporcional a la validez del test.
- 1.939 Rulon en "A simplified procedure for determining the reliability of a test by split-halves" presenta su famosa fórmula para el cálculo de la fiabilidad a partir de las puntuaciones de dos partes de un tests.
- 1.939 Frank introduce la expresión "técnicas proyectivas".

- 1.939 Mira y López publica su test miokinético de la personalidad.
- 1.941. Louis Guttman publica "The quantification of a class of attributes: A theory and method of scale construction" (En Horst, P. et al. The prediction of personal Adjustment. N.Y. Social Science Research Council), junto a sus trabajos de 1.944 y 1.950 constituye las bases de los procedimientos de escalación de respuesta categórica unidimensionales y determinísticos conocidos como "método de escalamiento de Guttman". Louis Guttman nació en 1.916 y murió en Minneapolis el 25 de Octubre de 1.987. Entre otros desempeño el cargo de director científico del "Israel Institute of Applied Social Research" y fue profesor de "medición social y psicológica" en la Universidad Hebrea de Jerusalén. Fue conocido por su "Facet Theory" y sobre todo por sus contribuciones al campo del escalamiento. Debe reseñarse que a Guttman le decepcionaba el modo en que comunmente se entendía su modelo de escalamiento acumulativo y los procedimientos de uso acrítico de su contribución.
- 1.941. S.S. Stevens presenta en el "International Congress for the Unity of Science" celebrado en septiembre, una clasificación de cuatro escalas de medida, nominal, ordinal, de intervalo y de razón, basada en las transformaciones admisibles.
- 1.941 Cyril Hoyt en un trabajo publicado en Psychometrica pone en relación el análisis de la fiabilidad y el análisis de varianza. Estos usos habían sido anticipados por R.W.B. Jackson en un trabajo publicado en el boletín del "Departament of Educational Research" de la Universidad de Toronto. Una nota sobre la fiabilidad publicada también en Psychometrica por Kaitz en 1.945 completan los trabajos pioneros sobre esta cuestión. En 1.978 Huck publicará una modificación sobre el procedimiento de Hoyt.
- 1.942 Hathaway y McKinley publican el "Minnesota Multiphasic Personality Inventory", que constituye el primer inventario clínico de evaluación de entidades gnosológicas psiquiátricas.
- 1.942 Robert W. B. Jackson estudia la relación entre "los viejos métodos de estimar la fiabilidad, que implican, excepto en el caso de pares-impares o dos mitades, dos aplicaciones del test" y "la nueva aproximación al problema de la medición de la fiabilidad sugerida por Kuder y Richardson (1.937) y por Hoyt (1.941)" que consiste "esencialmente en analizar los resultados de una única aplicación del test, y, bajo ciertos supuestos, obtener una estimación de la fiabilidad del test de los resultados de ese análisis." La principal conclusión del trabajo de Jackson es que, efectivamente, se trata de dos aproximaciones distintas que arrojan resultados diferentes dado que "no necesariamente se refieren a la misma cosa." Jackson establece que en los métodos de test-retest los errores de ambas aplicaciones no son independientes "posiblemente porque los individuos recuerdan los items y sus respuestas a ellos en el ensayo previo."
- 1.942 Truman L. Kelley publica un artículo titulado "El coeficiente de fiabilidad" donde revisa algunos de los principales conceptos y definiciones. A su parecer el coeficiente de fiabilidad implica en todo caso un juicio "a priori" de la equivalencia de medidas.

- 1.943 Harold Gulliksen publica en *Psychometrika* un trabajo titulado "A Course in the Theory of Mental Test" donde efectivamente presenta la estructura de un curso acerca de teoría de los tests, incluyendo un temario detallado y bibliografía, al modo de nuestros proyectos docentes actuales. A pesar de que el "curso" de Gulliksen es anterior a su propia obra de 1.950, a la de Lord y Novick de 1.968 y a todos los avances posteriores, el núcleo relativo a fiabilidad puede mantenerse parcialmente en la actualidad al tratar la teoría clásica de tests.
- 1.946 El Dr. S.S. Wilks en un trabajo titulado "Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution" se ponen las bases para poder establecer empíricamente si dos tests pueden ser vistos como paralelos en función de sus medias, varianzas y correlaciones. Esta es una cuestión principal para todo el trabajo práctico con el modelo de tests fundado en los trabajos de Spearman, el modelo de tests paralelos, donde el edificio teórico se construye bajo el supuesto de que se opera con tests paralelos. Wilks crea un estadístico llamado L_{mvc} , que varía entre cero y uno, adecuado para poner a prueba simultáneamente la hipótesis de que todas las medias, todas las varianzas y todas las correlaciones de tests supuestamente paralelos son iguales. Si medias, varianzas y covarianzas, son iguales su estadístico L_{mvc} vale 1. Wilks ofrece tablas para enjuiciar este estadístico con muestras pequeñas y sugiere como usar chi-cuadrado para su contraste en muestras grandes. Desde este estadístico se desprenden los casos para puesta a prueba de que, separadamente, medias, por un lado, o varianzas y covarianzas, por otro, son iguales.
- 1.947 E.A. Peel publica un método para calcular la fiabilidad máxima de una batería de tests. Al año siguiente publicará un trabajo acerca de la predicción de un criterio complejo considerando la fiabilidad de una batería de tests.
- 1.947 David Votaw, en su tesis doctoral, se ocupa de la igualdad de valideces que los tests paralelos deben mantener ante un criterio dado. En 1.948 Votaw presentará un test estadístico para igualdad de valideces de un conjunto de tests paralelos.
- 1.947 Primera edición de los tests de aptitudes diferenciales, DAT.
- 1.947 L.J. Cronbach publica en *Psychometrika*
- 1.947 El Employment Service publica su "General Attitude Test Battery".
- 1.947 Desde principios de siglo habían aparecido en los Estados Unidos diversos programas estatales o locales de tests, entre los que el más conocido fue el College Entrance Examination Board (CEEB). En 1.947 los servicios de pruebas del CEEB se fusionaron con los del American Council on Education y los de la Carnegie Corporation creándose un nuevo organismo llamado Educational Testing Service (E.T.S.). El Educational Testing Service ha cumplido desde ese momento un papel central en la elaboración de tests y pruebas por encargo de diversos tipos de instituciones y ha matenido una poderosa línea de investigación en el

campo, enriquecida con las aportaciones de muchas de las principales figuras americanas de la psicometría.

- 1.948. The Office of Strategic Services publica "Assessment of Men" donde incluye procedimientos multitests y técnicas situacionales y observacionales.
- 1.949 Publicación de dos manuales sobre tests psicológicos "Essentials of Psychological Testing" de L.J. Cronbach y "Mental testing" de F.L. Goodenough, ambos con especial énfasis en aspectos estadísticos.
- 1.949 L.J. Cronbach publica en el Psychological Bulletin una revisión de los trabajos que efectúan estudios con procedimientos estadísticos sobre las puntuaciones del test de Rorschach. La conclusión es que esos estudios han cometido serios errores estadísticos que vuelven sus resultados poco confiables.
- 1.949. Davis plantea una discusión acerca de la parcialidad de los tests contra los niños con nivel socioeconómico bajo.
- 1.950. Adorno publica su escala F.
- 1.950 Se publica una extensión de la escala Wechsler orientada para niños, la escala WISC, que consta de doce subtests que se agrupan en dos grandes escalas, una verbal y otra manipulativa.
- 1.950 Publicación del libro de S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, y J. A. Clausen, "Measurement and Prediction" (Princeton, N.J.; Princeton University Press) como volumen cuatro de una serie titulada "Studies in Social Psychology in World War II". En la obra hay contribuciones importantes de Guttman y también de Suchman sobre análisis de escalas. Los nueve primeros capítulos están escritos en su mayoría por Guttman y Suchman y constituyen una presentación y discusión del análisis de escalas asociado al modelo de escalograma de Guttman. El libro es importante sobre todo por la contribución de P.F. Lazarsfeld, que escribe los capítulos 10 ("The logical and mathematical foundation of latent structure analysis") y 11 ("The interpretation and computation of some latent structures"), realizando la primera exposición de la teoría de la estructura latente ("latent structure theory"). Lazarsfeld presenta la denominación de "ítem trace line" que se divulgará como denominación para la función que relaciona la probabilidad de una respuesta de un sujeto a un ítem, normalmente dicotómico, condicionada al nivel del rasgo latente que el sujeto posee. Se inicia con ello el análisis de los modelos de estructura latente en la versión que se extenderá dentro de la teoría de tests y escalamiento. El trabajo será continuado en un volumen publicado en 1.954. El mismo libro contiene el capítulo de Guttman "The basis for Scalogram analysis", una de las presentaciones más claras y populares del análisis de escalas tipo Guttman. En palabras de Guttman "El análisis de la escala pone a prueba la hipótesis de que un grupo de personas pueda ser ordenado en un orden jerárquico internamente significativo respecto a un área de datos cualitativos. Un ordenamiento jerárquico de personas es significativo si, a partir de la posición de la persona en el orden jerárquico, se conocen con precisión sus respuestas a cada uno de los ítems o enunciados incluidos en

la escala." Los capítulos finales (13 a 16) presentan diversos estudios concretos efectuados durante la guerra, siendo de un interés menos general.

- 1.950 "Theory of Mental tests" de Gulliksen. Se trata de una de las obras principales de la historia de la psicometría. Después de un primer capítulo introductorio, los capítulos dos y tres se ocupan de las ecuaciones que se desprenden de la definición de error aleatorio y de la definición de puntuación verdadera. El capítulo cuatro introduce los errores de medida, sustitución y predicción realizando una comparación entre los mismos que proseguirá en el capítulo siguiente al ocuparse de las varias interpretaciones del error de medida. Los capítulos seis, siete y ocho se ocupan de la relación entre longitud y fiabilidad. El capítulo nueve trata de la relación entre longitud y validez. El capítulo diez se ocupa del efecto de la variabilidad del grupo sobre la fiabilidad del test. El capítulo once del efecto de la variabilidad del grupo sobre la validez para el caso de dos variables. El capítulo doce de la selección univariada para el caso de tres variables, y el trece de la selección multivariada. El capítulo catorce recoge información sobre criterios estadísticos para afirmar que dos tests son paralelos. El capítulo quince lo titula métodos experimentales para obtener la fiabilidad de los tests y se ocupa principalmente de los métodos split-half. El capítulo dieciseis se ocupa de la estimación de la fiabilidad desde la homogeneidad de los items, recogiendo los desarrollos sobre las fórmulas del tipo KR-20. El capítulo diecisiete se ocupa de la cuestión de los tests de fondo versus los tests de velocidad desarrollando una serie de fórmulas originales para poner a prueba en que medida un test actúa como test de velocidad o como test de fondo. El dieciocho trata de los métodos de puntuación de los tests, restringiéndose sólo a los métodos de tiempo o de conteo de items y distinguiendo entre items con respuesta correcta e items sin respuesta correcta. El capítulo diecinueve de los métodos de estandarización e igualación (equating) de las puntuaciones. El veinte de la predicción diferencial y el veintiuno y último del análisis de items. En conjunto "Theory of Mental Tests", que ha sido recientemente reimpresso en 1.987, constituye una de las obras principales de la psicometría, presentando una visión integrada de una buena parte de los desarrollos en teoría clásica de tests producidos desde 1.904 hasta el mismo trabajo de Gulliksen en 1.950. El autor conoce los trabajos de Guttman, de Thurstone o de Lazarsfeld de ese mismo año y pone de manifiesto que la teoría clásica de los tests que expone en su obra se ha desarrollado al margen de la evolución de las teorías y métodos de la escalación nacidos desde la psicofísica, y que no integra propiamente ninguna de esas aproximaciones que exigirían sus propios métodos de análisis. Entre las aportaciones novedosas de la obra de Gulliksen pueden destacarse el estudio del efecto de la selección multivariada, la determinación de los componentes de velocidad y fondo, y una aproximación a la selección de items maximizando a la vez fiabilidad y validez. El mismo año, 1.950, Gulliksen publica también un artículo en *Psychometrica* acerca de la fiabilidad de los tests de velocidad.
- 1.950 La revisión de David A. Grant sobre teoría estadística y diseño de investigación dedica dos páginas a los métodos psicométricos, incluyendo técnicas psicofísicas, análisis de escalas y 'nuevas' aproximaciones al escalamiento. En estos apartados se pueden destacar las menciones de

los trabajos de Edwards sobre la técnica de escalamiento de Guttman, y los trabajos de escalamiento de Loevinger y de Coombs, ambos de 1.948. También se presentan bajo otro epígrafe desarrollos en análisis factorial, sin embargo, no se presta atención al campo de la teoría de tests.

- 1.950 Anne Anastasi, en un trabajo titulado 'El concepto de la validez en la interpretación de las puntuaciones de los tests', publicado en el *Educational and Psychological Measurement* (Vol. 10; pags. 67-78), realiza un análisis lógico del concepto de validez, destacando que solo cuando el test tiene propósitos de predicción o de diagnóstico la validez debe establecerse por un criterio externo. E.E. Cureton, en un trabajo publicado en el mismo volumen, pone de manifiesto con toda claridad que obtener el coeficiente de validez de los mismos datos que se usan para elaborar el test mediante los procedimientos de análisis de items es incorrecto. El coeficiente de validez obtenido de los mismos datos puede considerarse falaz.
- 1.951 Robert L. Thorndike presenta un capítulo titulado "Reliability" en un volumen editado por Lindquist. El capítulo hace una presentación global de la teoría y fórmulas de la fiabilidad sintetizando las principales aproximaciones clásicas.
- 1.951. S.S. Stevens edita el "Handbook of Experimental Psychology" en el que presenta un capítulo titulado "Mathematics, Measurement, and Psychophysics" donde expone de un modo sistemático su concepción de la medición, del lugar de la medición en la psicología, y sobre todo, sus famosas cuatro escalas de medida que ya había presentado en año 1.941.
- 1.951 J.L. Cronbach publica en *Psychometrika* "Coefficient alpha and the internal structure of tests."
- 1.951 El American Council of Education publica, siendo E.F. Lindquist el editor, la primera edición del handbook "Educational Measurement".
- 1.951 La revisión de Allen L. Edwards acerca de "Statistical theory and Research Design" incluye un apartado de apenas una página sobre psicofísica y escalamiento y otro ligeramente más extenso sobre tests psicológicos. Del apartado de escalamiento puede destacarse la mención del "método de dicotomías graduadas" desarrollado por Attneave en un trabajo del 1.949 y la del trabajo de Suchman de 1.950 repitiendo los argumentos del análisis de escalas contra el uso de las técnicas del análisis de items. Del apartado de test psicológicos pueden destacarse las referencias a los trabajos de Anastasi y Cronbach que hemos mencionado en los años inmediatamente anteriores. La revisión que cubre el periodo de junio de 1.949 a Mayo de 1.950 no alcanza a hacer mención de la obra de Gulliksen ni de la de Lazarsfeld.
- 1.952. Constitución de la Sociedad Española de Psicología.
- 1.952 Quinn McNemar orienta su revisión "Statistical theory and research methods" de modo mucho más próximo a los contenidos de la psicometría, dedicando por primera vez un espacio amplio a estos temas. Se dedica una extensa reseña, en tono de elogio, a la "Teoría de los tests Mentales",

de Gulliksen, que se considera "la publicación más importante aparecida en el campo en años" resaltándose su valor como sistematización y sus aportaciones originales.

- 1.953 Lindquist inicia el procedimiento de proceso electrónico de tests.
- 1.953 Técnica Q de Stephenson.
- 1.954 Se crea la Escuela de Psicología de la Universidad Central de Madrid, formalizándose los estudios de psicología.
- 1.954 Meehl publica "Clinical vs. statistical prediction".
- 1.954 La APA edita "Technical Recommendations for Tests" reeditado con revisión en 1.966 y en 1.974.
- 1.954 J.P. Guilford publica su "Psychometric Methods" (N.Y. MacGraw-Hill).
- 1.954 En la "International Conference on Testing Problems", cuyos trabajos fueron publicados al año siguiente, Saunders define el concepto de "moderator variable", variable moduladora.
- 1.955 Se publica la Wechsler Adult Intelligence Scale (WAIS) versión del test de Weschler-Bellevue con once subtests, seis pertenecientes a una escala verbal y otros cinco a una escala de ejecución.
- 1.955 Cronbach y Meehl publican "Construct validity in Psychological Tests". Este artículo retomaba el concepto de validez de constructo, inicialmente propuesto por P. E. Meehl y R.C. Challman y divulgado por las "Technical Recommendations for Psychological Tests and Diagnostic Techniques" publicadas como un suplemento del Psychological Bulletin en 1.954.
- 1.955 Ghiselli pone de manifiesto el bajo poder predictivo de muchos tests en "Differentiation of individuals in terms of their predictability" e inicia sus trabajos acerca de la predictibilidad, que están directamente relacionados con la cuestión de la mejora de la validez mediante técnicas de modulación.
- 1.956 Saunders plantea la cuestión de la modulación en ecuaciones de regresión.
- 1.956 Se celebra una sesión symposium acerca de la medición dentro del conjunto de sesiones de la "American Association for the Advancement of Science". Los trabajos presentados a esa mesa serán publicados tres años después por Churchman y Ratoosh, en un volumen con cuatro grandes partes: significado de la medición, teorías de la medida, problemas en las ciencias físicas y problemas en las ciencias sociales.
- 1.957 C.E. Osgood, G.J. Suci y P.H. Tannenbaum presentan "The measurement of Meaning" divulgando el diferencial semántico, que ya había sido presentado con anterioridad en 1.954. Para los autores el diferencial semántico "es un modo muy general de alcanzar un cierto tipo de

información, una técnica de medición generalizable en alto grado, que se debe adaptar a los requerimientos de cada problema de investigación al que se aplique." La obra presenta una exposición sistemática de los fundamentos, métodos de elaboración, de aplicación y de interpretación del diferencial semántico, ilustrada con ejemplos.

- 1.957 Cronbach pronuncia su famosa conferencia a la APA sobre "Las dos disciplinas de la Psicología Científica".
- 1.957 Cronbach y Gleser publican "Psychological Testing and Personnel Decision". La teoría de los tests se amplía hacia cuestiones de decisión.
- 1.957 Edwards presenta sus estudios sobre la deseabilidad social, poniendo de manifiesto la cuestión de los "estilos de respuesta". A.L. Edwards publica "Techniques of attitude scale construction" (N.Y. Appleton-Century-Crofts), un libro importante en la tradición del escalamiento de las actitudes.
- 1.958 Warren S. Torgerson publica "Theory and Methods of Scaling" uno de los libros más importantes publicados acerca del escalamiento psicológico. El libro contiene capítulos sobre teoría de la medida, clasificación de los métodos de escalamiento, distinguiendo entre los métodos de juicio y los métodos de respuesta, profundiza en los principales métodos de juicio relacionados con el escalamiento unidimensional, dedica un capítulo al escalamiento multidimensional, y dedica los últimos capítulos al escalamiento con métodos de respuesta. Se trata de una visión, ya clásica pero todavía en general vigente, ordenada y clarificadora del campo del escalamiento.
- 1.958 En Princeton tiene lugar un simposium sobre medición. Los principales trabajos serán publicados por Gulliksen y Messick en 1.960 bajo el título "Psychological Scaling: Theory and Applications".
- 1.959 Se queman públicamente miles de protocolos de tests en una escuela del estado de Texas. Se declaran ilegales los tests en el estado de California.
- 1.959 Campbell y Fiske formulan el método de la matriz multirasgo-multimétodo para el estudio de la validez convergente y discriminante. Esencialmente el procedimiento requiere la valoración de dos o más rasgos por dos o más métodos y el análisis de los coeficientes de correlación entre las puntuaciones obtenidas por los sujetos mediante los diversos métodos en los diversos rasgos.
- 1.959 Se establece el programa de examen American College Testing Program para aquellos centros que no utilizaban ya el más antiguo College Entrance Examination Board.
- 1.959 Guilford propone su famosa estructura tridimensional del intelecto que separa 120 habilidades específicas.
- 1.959 Se publica "The Measurement of Values" de Thurstone, obra que reunirá ordenadamente algunas de las principales contribuciones de Thurstone de otro modo dispersas en sucesivos artículos.

- 1.959 Obra de R.D. Luce "Individual Choice Behavior: A Theoretical Analysis."
- 1.959 Stevens añade a sus cuatro escalas originales la "escala de intervalo logarítmico" que después sería juzgada como en todo idéntica a la escala original de intervalos.
- 1.960 Tercera revisión del test de Stanford publicada por Terman y Merrill. El test presenta ahora una única escala (L-M) conteniendo los mejores elementos de la revisión de 1.937. Básicamente se realiza un trabajo de actualización y una nueva tipificación con 4.498 sujetos.
- 1.960 Se publica "Probabilistic Models for Some Intelligence and Attainment Tests" de G. Rasch. El modelo de Rasch es esencialmente muy simple. Según el modelo, la respuesta de un sujeto a un ítem (orientado a la medición de capacidad, aptitud o logro) depende de solo dos cosas: la aptitud del sujeto y la dificultad del ítem. Se trata de una las principales obras de la historia de la psicometría, abriendo un nuevo campo de trabajo que ha dado multitud de frutos posteriormente.
- 1.960 Con motivo del centenario de los Elementos de Psicofísica de Fechner y del veinticinco aniversario de la fundación de la Sociedad Psicométrica, esta última organiza una serie de sesiones especiales. Entre ellas el Simposium "Psicofísica: Cien años después". Los principales trabajos serían publicados en *Psychometrika* en 1.961.
- 1.961 Primera revisión en el "Annual Review of Psychology" bajo la denominación de "Escalamiento y teoría de tests", encargada a Warren S. Torgerson, del "Massachusetts Institute of Technology". Antes de esta fecha no hay revisiones en el "Annual" de los temas escalamiento o teoría de tests, sin embargo, en las revisiones anteriores acerca de estadística o métodos estadísticos es posible encontrar alguna mención a estos temas. La revisión de Torgerson se ocupará de los siguientes temas: la lógica de la medida; el escalamiento psicofísico; la discriminabilidad; el escalamiento multidimensional; el diferencial semántico; juego, utilidad y probabilidad subjetiva; teoría de tests; y análisis factorial. En consonancia con la trayectoria del autor, que hacía solo tres años había dado a luz una de las principales obras sobre el escalamiento, la revisión puede considerarse muy desequilibrada del lado del escalamiento, con una atención mucho más limitada a la teoría de los tests.
- 1.962 Este año se considera el punto de partida de la "medición referida-al-criterio". Glaser y Klaus presentan este año un trabajo que utiliza por vez primera el término "criterion-referenced-test". Sin embargo, el trabajo más divulgado como iniciador de esta orientación de trabajo es el de Glaser de 1.963 "Instructional technology and the measurement of learning outcomes". En este trabajo Glaser ponía de manifiesto la existencia de un continuo subyacente de capacidad del sujeto, desde la ineficacia hasta el perfecto desempeño, y señalaba las medidas referidas al criterio como aquellas que identificaban algún punto de corte o standard que el sujeto debe superar para ser considerado 'master' en la materia. Los trabajos de Mager de 1.962, de Popham y Husek de 1.969, de Iven de 1.970, de Glaser y Nitko de 1.971, de Lidvall y Nitko de 1.975 contribuyeron a la formación de los

conceptos básicos en el campo. Los de Ebel de 1.972, Nedelsky de 1.954, y Angoff de 1.971 contribuyeron especialmente a dotar de técnicas para el establecimiento de criterios. Técnicas que fueron discutidas en un interesante trabajo de Gasll de 1.978 y que, junto con los principales desarrollos teóricos de la materia, pueden verse estructuradamente a partir de las sucesivas revisiones sobre el tema encabezadas por Hambleton y publicadas en el *Review of Educational Research* y en el *Journal of Educational Measurement* principalmente. Debe observarse que la medición referida al criterio presenta una estrecha conexión con los modelos de estructura latente, con el análisis de clase latente si se supone que los sujetos son de dos tipos posibles (master o no master) y con la teoría de la respuesta al ítem si se supone un continuo de capacitación en el dominio.

- 1.962 Publicación de los primeros trabajos de Shepard sobre escalamiento multidimensional no métrico, que junto con los de Kruskal de 1.964 sobre el mismo tema, pueden considerarse el arranque del gran desarrollo reciente del escalamiento multidimensional.
- 1.963 Glaser propone los "Criterion-referenced tests" frente a los "Norm-referenced tests".
- 1.963 Aparece en el 'Annual Review of Psychology' la segunda revisión del campo "escalamiento y teoría de los tests". Esta vez el encargado de la misma es Ledyard R. Tucker, de la Universidad de Illinois. La revisión cubrirá los tres grandes apartados: escalamiento psicológico, teoría de los tests, y análisis factorial. En esta revisión el espacio dedicado a la teoría de los tests es comparable al utilizado en cada uno de los otros tópicos. Es la última revisión que considerará conjuntamente el escalamiento y la teoría de los tests, en siguientes revisiones del A.R.P. estos campos serán tratados separadamente.
- 1.963 L.J. Cronbach, N. Rajaratnan y G. Glesser publican "Theory of generalizability: a liberalization of reliability theory" lo que constituye la primera presentación formal de la teoría de la generalizabilidad, ampliada después en un trabajo de 1.972 de Cronbach, Gleser, Nanda y Rajaratnam.
- 1.963 Desde principios de siglo han aparecido alusiones a métodos de clasificación, sin embargo, la obra de dos biólogos Robert Sokal y Peter Sneath, "Principles of Numerical Taxonomy" publicada este año abre definitivamente el camino al desarrollo de los métodos de análisis Cluster. La mayoría de la literatura sobre cluster puede encontrarse desde esta fecha hasta nuestros días. La idea central de estos autores consiste en tomar un conjunto de datos relevantes acerca del conjunto de organismos que constituyen el objeto del estudio, obtener una estimación del grado de similitud entre ellos y utilizar un método cluster para clasificar los organismos en grupos en función de su similitud.
- 1.964 Se prohíbe la aplicación de tests en las escuelas de Nueva York por orden judicial.

- 1.964 C.H. Coombs concluye un vasto programa de investigación de los modelos de escalamiento no métricos con la publicación de "A Theory of data" (N.Y. Wiley).
- 1.964 N. Frederiksen y H. Gulliksen publican sus "Contributions to Mathematical Psychology" donde se incluye una extensa bibliografía y una biografía de Thurstone.
- 1.965 Dos comités del Congreso de EEUU estudian el tema de los tests y se acaba prohibiendo el uso de los tests de personalidad a la hora de seleccionar empleados federales. En el contexto del debate social sobre los tests y su uso, la APA nombra un comité especial para el estudio de las dimensiones sociales del uso de los tests y publica un número monográfico del *American Psychologist* sobre el tema.
- 1.965 Zubin, Eron y Schumer publican "An Experimental Approach to Projective Techniques".
- 1.965 Primera revisión en el 'Annual Review of Psychology' del campo 'escalamiento' tomado separadamente de la teoría de tests. La revisión la realizan Cösta Ekman y Lennart Sjöberj de la Universidad de Estocolmo. Presenta seis grandes apartados: problemas generales, métodos thurstonianos, métodos de estimación, estudios comparativos, métodos multidimensionales, y métodos de escalamiento y la ley psicofísica. La revisión inmediatamente precedente de los desarrollos sobre escalamiento debe buscarse fuera del 'Annual', en el trabajo de Schiffman y Messick "Scaling and measurement theory: a brief review of developments from 1.960 to 1.963" apareció en 1.963 (*Rev. Educ. Res.* Vol. 33, pags 533-42.)
- 1.965 L.J. Cronbach y G.C. Gleser publican "Psychological tests and personnel decisions" (Urbana, University of Illinois Press).
- 1.966 Ghiselli revisa toda una serie de investigaciones sobre validez encontrando que un coeficiente de validez raramente sobrepasa .50, lo cual sugiere la necesidad de encontrar alternativas a los modelos clásicos de validación.
- 1.966 La American Psychological Association publica una edición de su "Standars for Educational and Psychological Tests and Manuals" prescribiendo normas para la elaboración de los tests.
- 1.966 Se edita en EEUU una traducción inglesa de los "Elementos de Psicofísica" de G. Fechner.
- 1.967 Publicación de "Quantification in Psychology" de W.L. Hays y de "Psychometric Theory" de J.C. Nunnally. Se trata de dos textos de carácter más bien introductorio y orientados principalmente a la enseñanza para pregraduados. El texto de Nunnally puede encontrarse en una versión castellana editada en México D.F. por Trillas en 1.987. La obra se divide en cuatro grandes partes: los principios básicos, la estructura interna de las medidas, el análisis multivariado y las áreas de contenido. En medio de una exposición no siempre ordenada, pueden encontrarse algunas aportaciones e intuiciones sugestivas sobre teoría de tests. En 1.978 Jum

C. Nunnally presentará una segunda edición inglesa revisada. Jum Clarence Nunnally nació en Birmingham (Alabama), en 1.924 y falleció en 1.982 en Nashville (Tennessee) la obra "Psychometric Theory" que presentó por primera vez en 1.967 constituye su contribución más visible al campo de la psicometría, sin embargo también debe destacarse su labor institucional en la dirección del departamento en la Universidad de Vanderbilt del 61 al 64 y del 67 al 70, y sus innumerables contribuciones como revisor de varias revistas.

- 1.967 Aparece la primera revisión titulada "Teoría de tests" en el 'Annual Review'. En 1.961 y en 1.963 el campo había sido revisado, pero solo como un apartado en artículos que incluían además y sobre todo escalamiento. En esta ocasión J. A. Keats, de la Universidad de Newcastle (Australia), es el encargado de la revisión. Keats presenta su revisión bajo seis grandes apartados: ordenación de sujetos, ordenación conjunta de sujetos e items, medición conjunta, fiabilidad, análisis de items, y puntuación de los tests, añadiendo unas sugerencias finales para la investigación futura.
- 1.967 D. Magnusson publica su obra "Test theory" (Reading MA. Addison-Wesley).
- 1.968 P.F. Lazarsfeld y N.W. Henry publican "Latent Structure Analysis" donde se describen los diversos modelos de estructura latente que pueden diseñarse, analizándose los aspectos estadísticos de los mismos. Entre otras conexiones se describe, por ejemplo, la relación del escalamiento de Guttman con el modelo de clase latente.
- 1.968 Anne Anastasi publica su "Tests Psicológicos", otro texto de carácter predominantemente docente, más orientado a presentar una panorámica de los tests psicológicos disponibles y sus condiciones de enjuiciamiento con una óptica profesional práctica que a adentrarse en la teoría psicométrica de los tests psicológicos. El libro, además de dedicar una parte a los aspectos de fundamentación de los tests psicológicos (fiabilidad, validez, análisis de elementos), y otra a una perspectiva de la situación de los tests en su época, se ocupa de los tests de desarrollo intelectual general, de los tests de aptitudes diferenciales y de los tests de personalidad. Entre los tests de personalidad incluye los inventarios autodescriptivos, las medidas de intereses y actitudes y también las técnicas proyectivas.
- 1.968 F.M. Lord y M.R. Novick, con contribuciones de Birnbaum, publican la que puede considerarse obra central de la psicometría en el campo de la teoría de tests: "Statistical Theories of Mental Test Scores". El libro está dedicado a la teoría de los tests, pudiendo considerarse la culminación del desarrollo de las teorías clásicas de los tests, incluyendo la teoría de la Respuesta al Item, pero también presenta algunas aportaciones sobre escalamiento. Ha sido considerado como el hecho de "mayor impacto en el campo" (Lewis, 1.986), "Un gran paso hacia adelante" (Bock y Wood, 1.971) y también "un gran mausoleo" (Lumsden, 1.976) por algunas de las principales revisiones en el campo.
- 1.968 Se publica el trabajo de Mischel "Personalidad y Evaluación" que representa un crítica dura a la concepción de medición de rasgos mediante tests.

- 1.968 Aparece el libro de R.D. Bock y L.V. Jones "The measurement and prediction of judgement and choice" que ha sido considerado la continuación del trabajo de Torgerson "Theory and methods of scaling" aparecido diez años antes. Sin embargo, se concentra solo en los modelos thurstonianos y no se ocupa del escalamiento multidimensional.
- 1.968 McReynolds edita el primer volumen de "Advances in Psychological Assessment".
- 1.968 Harold Gulliksen publica un artículo titulado "Métodos para determinar la equivalencia de medidas" que presenta una interesante discusión de la cuestión de la medición de variables latentes y la delimitación de qué miden diversos indicadores o tests de una misma variable o constructo psicológico. Entre otros temas el artículo discute lo que denomina el punto de vista de "Humpty Dumpty" (personaje de Carroll en "A través del espejo" que definía por sí solo sus palabras) en la medición.
- 1.969 La APA edita indicaciones para el examen de personas poco dotadas.
- 1.969 Butcher realiza interpretaciones automatizadas de cuestionarios MMPI.
- 1.969 Revisión de Joseph L. Zinnes del tópico "Scaling" en el "Annual Review of Psychology". Mientras que la revisión anterior del campo por Ekman y Sjöberg, realizada en 1.965, criticaba los conceptos de la teoría de la elección, esta revisión tomará el concepto de elección como eje de su trabajo. La revisión cubrirá los siguientes temas: teorías y experimentos de comparación de pares, lógica de la medida, medida conjunta, modelos multidimensionales, funciones potenciales versus logarítmicas, y detección.
- 1.969 Benjamin Wright y Nargis Panchapakesan en el trabajo "A procedure for sample-free item analysis" describen un procedimiento para realizar un nuevo tipo de análisis de ítems. El procedimiento se funda en el modelo de Rasch y presenta dos características principales. Primero, que las calibraciones del test son independientes de la muestra de sujetos utilizada para la estimación de los parámetros de los ítems. Segundo, que las mediciones de las personas son independientes de la selección de ítems utilizada para obtener las puntuaciones de los tests.
- 1.970 L.J. Cronbach presenta la tercera edición de su "Essentials of Psychological testing" (N.Y. Harper and Row).
- 1.971 Sheldon Zedeck publica "Problems with the use of Moderator Variables" donde presenta una síntesis de los principales métodos de modulación y algunas de las cuestiones críticas más importantes respecto al tema que vienen arrastrándose desde que en 1.928, y por primera vez en la historia de la validación y predicción en psicología, Hull hiciera notar con inquietud que el límite superior empírico de los índices de validez estaba en torno a .50.
- 1.971 Segunda edición de "Educational measurement" editado por R.L. Thorndike. La contribución de R. Gleser y A.J. Nitko titulada "measurement in learning and instruction" adelanta una de las más citadas definiciones de tests

referidos al criterio: "Un test referido al criterio es aquel deliberadamente construido para arrojar medidas que son directamente interpretables en términos de standards de desempeño específicos." La contribución de W.H. Angoff titulada "Scales, norms, and equivalent scores" puede considerarse el primer trabajo con un tratamiento extensivo de la cuestión de la 'igualación' de tests ('equating') que se desarrollaría rápidamente desde los primeros años ochenta.

- 1.971 Revisión del campo de la "teoría de los tests" por R. Darrell Bock y Robert Wood, ambos del Departamento de Educación de la Universidad de Chicago. Se pasa revista a uno de los periodos más fructíferos del desarrollo y, sobre todo, de la sistematización de la teoría de los tests. En los cuatro años transcurridos desde la última revisión, por Keats en 1.969, han aparecido una serie de obras relevantes entre las que los revisores han de destacar claramente la que llaman el "suceso señal del periodo", el trabajo de Lord y Novick de 1.968 al que ya nos hemos referido. La misma estructura de la revisión de Bock y Wood está influida por el trabajo de Lord y Novick. La revisión ordena sus contenidos en torno a los siguientes grandes núcleos: teoría general de los tests, teoría clásica de los tests, teoría fuerte de las puntuaciones verdaderas, teoría del rasgo latente, otros tópicos, y, por primera vez, procedimientos informáticos. Destaca la atención que se presta a la teoría del rasgo latente y que algunos contenidos como "fiabilidad" a los que con anterioridad se dedicaba una atención preferente sean relegados ahora al lugar de subapartados.
- 1.972 J.L. Cronbach, G.C. Gleser, A.N. Nanda, y N. Rajaratnam publican "The Dependability of Behavioural Measurements: Theory of generalizability for scores and profiles", una obra clave en la teoría de la generalizabilidad.
- 1.972 R. Shepard, A.K. Romney y S. Nervlone editan " Multidimensional Scaling: Theory and applications in the behavioral sciences" (N.Y. Seminar Press) obra en dos volúmenes (el segundo es firmado primero por Romney) que constituye una exposición sistemática de las técnicas del MDS.
- 1.973 Sneath y Sokal clasifican los coeficientes de similitud en cuatro grupos principales: coeficientes de correlación, medidas de distancias, coeficientes de asociación y medidas de similitud probabilísticas.
- 1.973 Revisión del campo "escalamiento" por Norman Cliff. La revisión comienza con la declaración de principios de que "el escalamiento o la medición no es simplemente la asignación de números a eventos según reglas". Se trata de una revisión que realiza una detallada clasificación de las aportaciones en campos y subcampos. Los principales tópicos considerados son: teoría axiomática de la medición, escalamiento con modelos funcionales, escalamiento multidimensional, escalamiento unidimensional, y relaciones psicofísicas.
- 1.974 La "American Educational Research Association" y el "National Council on Measurement in Education " de la American Psychological Association publican una edición de sus "Standards for Educational and Psychological Tests."

- 1.974 "Measurement in the social sciences: Theories y strategies" de H.M. Blalock.
- 1.975 Cronbach presenta "Beyond the two disciplines of Scientific Psychology."
- 1.976 La obra dirigida por Pawlik "Diagnose der Diagnostik" plantea una serie de críticas replanteando el diagnóstico mediante tests.
- 1.976 En el 'Annual Review of Psychology' se publica la revisión de James Lumsden, de la Universidad de Western Australia, sobre el tema "teoría de tests". Lo esencial (y lo original) del trabajo de Lumsden es que se trata de una revisión crítica. Una revisión que es más bien una reflexión sincera, desencantada y en ocasiones ácidamente crítica con los desarrollos alcanzados por la teoría de los tests hasta el momento. Una serie de cuestiones como las puntuaciones verdaderas y la fiabilidad son sometidas a intensa revisión sugiriéndose en unos casos el abandono de antiguos conceptos y en otros su reinterpretación. Desde luego Lumsden no propone abandonar el trabajo en teoría de tests, sin embargo, sugiere encarecidamente la necesidad de una renovación.
- 1.976 R.L. Wood, M.S. Wingersky, y F.M. Lord presentan el programa LOGIST orientado al cálculo de los parámetros de las c.c.i. y de la capacidad del sujeto según la TRI. El programa LOGIST nació en el seno del Educational Testing Service (Princeton, NJ) y es quizás el programa de estimación de parámetros según la TRI que más popularidad ha alcanzado en los años sucesivos. La guía de usuario dada a la luz ese mismo año aparece firmada solo por el primer y el último autor reseñados.
- 1.977 Hambleton y Cook presentan una introducción a la teoría de la respuesta al ítem en un artículo publicado en el Journal of Educational Measurement (Vol. 14) titulado "Latent trait models and their use in the analysis of educational test data".
- 1.978 John C. Baird y Elliot Noma publican "Fundamentals of Scaling and Psychophysics" (N.Y. Wiley), un libro que esta enfocado a tratar el enlace clásico entre psicofísica y escalamiento. Los capítulos uno a siete y el capítulo nueve tratan acerca de la psicofísica desde un punto de vista clásico, el resto de los capítulos enfrentan desde un punto de vista de escalamiento un conjunto de temas relacionados: la teoría de la detección de señales, el escalamiento multidimensional, el análisis cluster, y la teoría de la información.
- 1.978 Segunda edición revisada de la "Psychometric Theory" de Jum C. Nunnally.
- 1.978 Joseph B. Kruskal y Myron Wish publican "Multidimensional Scaling" (en la serie "Quantitative Applications in the Social Sciences" de Sage). Se trata de un volumen introductorio que puede resultar muy útil en un primer curso sobre escalamiento. Efectúa una extensa introducción a los conceptos básicos del escalamiento multidimensional, de una gran claridad didáctica, una discusión del tema de la dimensionalidad y el ajuste y una presentación del método INDSCAL.

- 1.978 Samuel Shye edita "Theory Construction and Data Analysis in the Behavioral Sciences" que se ocupa de las aportaciones a la teoría de medición de Louis Guttman, con discusión de algunos tópicos avanzados. El libro discute por ejemplo el "Smallest Space Analysis" el "Multidimensional Scalogram Analysis" y el "Partial Order Scalogram Analysis."
- 1.979 John L. Sullivan y Stanley Feldman publican "Multiple Indicators. An Introduction". El libro concierne el tema del uso de indicadores múltiples de los conceptos teóricos y se ocupa básicamente de dos grandes contenidos: las matrices multirasgo-multimétodo y, propiamente, los indicadores múltiples. El volumen implica un cierto conocimiento de las técnicas de modelos causales en diseños correlacionales y de la teoría de la fiabilidad y la validez.
- 1.979 Edward G. Carmines y Richard A. Zeller publican "Reliability and Validity Assessment" (en la serie "Quantitative Applications in the Social Sciences" de Sage). El volumen trata en menos de cincuenta páginas sobre los principales métodos prácticos de cálculo de la fiabilidad y la validez. Se trata de una obra de lectura extraordinariamente sencilla que puede recomendarse como introducción en un curso muy elemental sobre estos temas. Sin embargo, precisamente por esta orientación no puede considerarse adecuado ni siquiera como un manual suficiente para un curso introductorio, pudiendo considerarse una lectura complementaria o una presentación sumaria para los primeros pasos acerca de estos temas.
- 1.979 Fred S. Roberts es el autor del volumen siete de la "Encyclopedia of Mathematics and its Applications", titulado "Measurement Theory" que constituye una introducción a la medición enfocada a las ciencias sociales con una orientación representacional. El libro puede considerarse una pieza interesante dentro del campo interdisciplinar de la teoría de la medición, desde un enfoque de clarificación y fundamentación matemática. En este campo y con este mismo nivel introductorio, orientados hacia la fundamentación matemática de la teoría de la decisión, pueden mencionarse las obras de Raiffa de 1.968 "Decision Analysis" y de Keeney y Raiffa de 1.976, ambas con influencias de la anterior de Fishburn "Utility theory for decision making" de 1.970. A un nivel más complejo dentro del campo de fundamentación matemática de la medición y con una amplitud más comprehensiva se mueven los trabajos de Pfanzagl "Theory of Measurement" (N.Y.; Wiley) de 1.968 y de Krantz, Luce, Suppes, y Tversky "Foundations of Measurement" (N.Y.; Academic Press) de 1.971. El trabajo de Suppes y Zinnes de 1.963 "Basic Measurement Theory" incluido en el "Handbook of Mathematical Psychology" de Luce, Bush y Galanter, que extendió las aportaciones de el de Scott y Suppes "Foundational Aspects of Theories of Measurement" (J. of Symbolic Logic. Vol. 23) de 1.958 están también en la base de los desarrollos del campo.
- 1.979 Levine y Rubin publican un trabajo titulado "Measuring the appropriateness of multiple-choice test scores" que puede ser considerado uno de los pioneros en el tratamiento de la cuestión de adecuación, sugiriendo un índice de adecuación.

- 1.979 M.J. Allen y W.M. Yen publican "Introduction to Measurement Theory", un libro que constituye un manual introductorio a la psicometría. El capítulo uno contiene una breve presentación histórica, y el dos una revisión de los conceptos estadísticos básicos. El capítulo tres trata de la teoría clásica de las puntuaciones verdaderas. Los capítulos cuatro y cinco se ocupan de la fiabilidad y la validez. Los capítulos seis, siete y ocho del análisis de ítems, la construcción de tests, los métodos de igualación y varios métodos de escalamiento. El capítulo nueve se ocupa de una variedad de tópicos incluyendo regresión múltiple, heterogeneidad del grupo, sesgos de selección, análisis de perfiles, y medida del cambio. El capítulo diez trata sobre una serie de controversias y aspectos críticos. Por último, el capítulo final trata sobre la teoría fuerte de las puntuaciones verdaderas.
- 1.980 Arthur R. Jensen publica "Bias in mental Testing". El interés de Jensen por este tema aparece en torno a 1.950, y proviene de Kenneth Eells, el que fuera su maestro y director de su tesis doctoral y uno de los mayores especialistas en el tema de los sesgos debidos a clase social. Jensen no obstante extendió su preocupación a sesgos cuyo origen no fuera la posición social. La obra se ocupa del tema de los sesgos de los tests mentales en función de las condiciones personales, raciales o sociales de los sujetos. Jensen realizó una revisión exhaustiva de la investigación sobre sesgos. Junto a este tema central la obra articula una interesante discusión sobre las críticas recibidas por los tests mentales y los abusos en la utilización de los tests. La principal (y polémica) conclusión de la obra es que los más importantes tests de capacidad mental expresados en inglés no están sesgados respecto de minorías anglo-parlantes en los Estados Unidos y que, para los tests más importantes no verbales, esta afirmación de la ausencia de sesgo puede extenderse a los grupos no anglo-parlantes. En 1.969 Jensen había publicado una pequeña monografía "How Much Can We Boost IQ and Scholastic Achievement" en la que se argumentaba que la educación compensatoria había sido un fracaso en buena parte por una inadecuada concepción de la inteligencia humana. Jensen argumentaba allí contra la hipótesis del niño medio, que supone que existen pocas diferencias en la potencialidad intelectual, y contra la hipótesis de la privación social, que pretende explicar las diferencias individuales y sociales en IQ (CI). Por el contrario Jensen sostiene el punto de vista que, apoyado por datos, ya defendiera Terman en 1.916, en el sentido de que existen fuertes diferencias en las capacidades humanas. Frente al trabajo de Jensen se alzaron una serie de voces, entre ellas la del francés Kamin en 1.974, defendiendo la ausencia de evidencia para la heredabilidad de las puntuaciones de los tests de inteligencia. Jensen ha considerado en los tres primeros capítulos de su obra de "Bias in mental testing" el conjunto de críticas y el movimiento social contra el IQ. El capítulo seis se ocupa de la relación entre inteligencia e IQ. De especial interés psicométrico es el capítulo nueve acerca de la definición de sesgo y de test sesgado. En resumen, según Jensen, un test puede considerarse sesgado si las mismas puntuaciones en el test predicen diferentes niveles de desempeño en el grupo minoritario, si el contenido de algunos ítems favorece algún grupo, y si aparecen diferencias de grupo debido a factores no intrínsecos al test tales como características del examinador, diferencias en experiencia con tests, factores motivacionales o de personalidad. En los capítulos siguientes se ocupa de revisar la evidencia acerca del sesgo de

los test mentales y por último, en el capítulo quince se ocupa de los abusos de los test mentales.

- 1.980 Con motivo de su cincuenta aniversario el Australian Council for Educational Research celebra un seminario internacional, que cuenta con la participación de R.L. Thorndike, titulado "The Improvement of Measurement in Education and Psychology" que fue orientado hacia las aportaciones de las teorías del rasgo latente. Las actas fueron editadas por Donald Spearrit en 1.982 y subtituladas "Contributions of Latent Trait Theories". El libro contiene una revisión de R. L. Thorndike sobre medición en educación y una serie de capítulos acerca del uso de la Teoría de la Respuesta al Ítem en la medición en educación. Entre ellos puede mencionarse el John A. Keats que establece una comparación entre la teoría clásica de tests y la teoría del rasgo latente.
- 1.980 Erling B. Andersen publica "Discrete Statistical Models with Social Science Applications". El libro merece ser citado aquí porque presenta un conjunto de desarrollos sobre los modelos logísticos que están a la base de una serie de procedimientos de calibración de ítems y estimación de parámetros en la TRI.
- 1.980 F.M. Lord publica "Applications of Item Response Theory to Practical Testing Problems", una obra que puede considerarse central en el desarrollo de la teoría de la respuesta al ítem. El capítulo uno ofrece un sintético sumario de la teoría clásica de tests. El capítulo dos es una introducción a la teoría de la respuesta al ítem, y el tres una comparación de la teoría de la respuesta al ítem y de la teoría clásica. Los capítulos cuatro y cinco profundizan en la teoría de la respuesta al ítem. El capítulo cuatro se ocupa básicamente de la estimación de parámetros, considerando las puntuaciones de los tests en función de esos parámetros. El capítulo cinco trata acerca de las funciones de información y de los pesos óptimos de ponderación de ítems. Con este capítulo termina la primera parte del libro dedicada a una introducción de la teoría de la respuesta al ítem. La parte segunda se ocupa de un conjunto de aplicaciones de la teoría de la respuesta al ítem, incluyendo los tests 'flexilevel', los tests 'multilevel', los 'tailored' y los 'mastery' tests. La parte tres del libro se dedica a problemas prácticos y otras aplicaciones. El capítulo doce se ocupa de profundizar en el tema de la estimación de la capacidad y de los parámetros, que ya había sido introducida en la primera parte. También se trata el tema de la igualación, del sesgo de los ítems y de las respuestas omitidas. Por último, la parte cuarta del libro se ocupa de la estimación de la distribución de las puntuaciones verdaderas.
- 1.980 Gaito publica en el Psychological Bulletin "Measurement, scales and statistics: resurgence of an old misconception" acerca de la vieja cuestión, que fue inaugurada por Stevens en 1.951, sobre la relación entre las escalas de medida y los estadísticos que cada una de las escalas puede soportar.
- 1.980 J. Douglas Carroll y Phipps Arabia publican la primera revisión en el 'Annual Review of Psychology' encabezada "escalamiento multidimensional." Hasta este momento el escalamiento multidimensional era solo un apartado

dentro de las revisiones de escalamiento. La revisión se hace eco de los desarrollos acaecidos desde la de Cliff en 1.973 y clasifica las aportaciones usando una nueva taxonomía que intenta ordenar el campo. En realidad su taxonomía es una revisión y actualización, con aportaciones originales, de la que Coombs propuso en "A theory of data" en 1.964. El mismo número del A.R. of P. contiene la revisión de Bentler "Multivariable analysis with latent variables: causal modeling.", en el, por varias razones, campo próximo de los modelos causales.

- 1.980 L.J. Th. Van der Kamp, W.F. Langerak, y D.N.M. de Gruijter editan el volumen titulado "Psychometrics for Educational Debates" (Chichester, Wiley). La obra recoge una selección de los trabajos presentados en el "Third International Symposium on Educational Testing" celebrado en Leyden en 1.977. Se trata de una obra heterogénea e irregular donde junto a trabajos de H.J. Eysenck sobre la heredabilidad de la inteligencia se cuentan otros de carácter propiamente psicométrico sobre TRI, sesgos o 'tailored testing'. Entre los autores se cuentan M.R. Novick, F.M. Lord, y los mismos editores.
- 1.980 R. A. Berk es el editor de "Criterion-Referenced Measurement: The State of Art" (Baltimore. The Johns Hopkins University Press.) que recoge seis trabajos presentados en el primer anual "Johns Hopkins University National Symposium on Educational Research" que se celebró en 1.978 en Washington D.C. El volumen repasa algunos de los principales temas sobre tests referidos al criterio.
- 1.980 Richard A. Zeller y Edward G. Carmines publican "Measurement in the Social Sciences", subtítulo "The link between theory and data". El libro tiene el propósito de realizar una introducción a la medición en el campo de las ciencias sociales atendiendo a su justificación y procedimientos, con una especial atención a la consideración del error sistemático de medida. El libro presenta seis capítulos y un apéndice. El primero efectúa una introducción a la medición donde los autores rechazan la concepción de la medición de Stevens, y partiendo de una sugerencia de Blalock, redefinen la medición como el proceso de enlazar conceptos abstractos a indicadores empíricos. En el mismo capítulo hay una presentación sumaria de la teoría clásica de las puntuaciones verdaderas, siguiendo de cerca a Lord y Novick, y una reformulación de la teoría clásica introduciendo en las ecuaciones el error sistemático de medida. El capítulo dos hace una presentación conceptual e intuitiva, no matemática del análisis factorial. Los capítulos tres y cuatro se dedican respectivamente a la fiabilidad y a la validez. Estos capítulos presentan un contenido semejante al libro de los mismos autores "Reliability and Validity Assessment" de 1.979. El capítulo cinco se dedica íntegramente a la evaluación del error sistemático de medida. El capítulo seis se dedica a integrar los conceptos de fiabilidad y validez, y contiene además una revisión y una síntesis de las principales aportaciones de la obra. El apéndice, firmado también por McIver, (el coautor junto a Carmines de la obra "Unidimensional Scaling" también de 1.979), se ocupa del tema de los indicadores múltiples, ocupándose de una aplicación del método path y otra del método de estructuras de covarianza para identificar error sistemático.

- 1.980 Richard Bell y James Lumsden publican un trabajo titulado "Test Length and validity" donde examinan empíricamente la relación entre la longitud del test y la validez predictiva. La principal conclusión del trabajo es que los tests de capacidad analizados (cuatro) pueden ser reducidos en un 60% sin un decremento considerable en la validez. Los autores sugieren que un estudio de la relación validez-longitud debe llevarse a cabo para los principales tests de capacidad hasta lograr la longitud óptima.
- 1.980 Reimpresión de la obra de Rasch, original de 1.960, "Probabilistic Models for Some Intelligence and Attainment Tests". Se incluye ahora un "prólogo" y un "epílogo" por Benjamin D. Wright. Wright no duda en calificar la obra de Rasch como "el trabajo más importante en psicometría desde los artículos de Thurstone de los años 1925 a 1929 y la monografía de 1929 por Thurstone y Chave. La investigación psicométrica realizada por Rasch entre 1951 y 1959, que él explica e ilustra en este libro, marca el punto en que la psicometría pasa de ser puramente descriptiva a convertirse en una ciencia de medición objetiva"
- 1.980 Ronald K. Hambleton edita "Contributions to Criterion-Referenced Testing Technology".
- 1.980 Van der Ven publica su "Introduction to Scaling" (N.Y. Wiley), un libro que puede ser útil para docencia con estudiantes con algún curso previo de estadística. El libro trata una serie extensa de tópicos incluyendo escalamiento de Guttman, escalamiento de despliegamiento de Coombs, escalamiento multidimensional no métrico, etc. con una atención preferente a los modelos deterministas y una atención limitada a los probabilísticos. El libro puede ubicarse en la tradición de Guilford (1.954), de Torgerson (1.958) y de Coombs (1.964), excluyendo la cuestión de la medición de las actitudes.
- 1.981 Aparece el libro de E.E. Ghiselli, J.P. Campbell y S. Zedeck "Measurement Theory for the Behavioral Sciences", una de las exposiciones más intuitivas, didácticas y esclarecedoras de los aspectos generales y básicos de la teoría de la medición y su conexión con la teoría de la correlación y regresión. El manual recorre como principales temas: la descripción de las diferencias individuales, los aspectos básicos de la psicología de la medición, transformación de puntuaciones, el concepto de correlación y el de regresión, trata la fiabilidad, la validez, predicción y los conceptos básicos del escalamiento psicológico. Es un libro de orientación docente para un curso introductorio que expone con frescura un conjunto seleccionado y limitado de tópicos. Sin embargo, la mayoría de los desarrollos recientes y de las complejidades clásicas de la teoría de la medición y de la teoría de los tests no son tratados en absoluto.
- 1.981 "Ethical principles for psychologists" de la APA, publicados en *American Psychologist* (Vol 36. Pags 633-638).
- 1.981 Gideon J. Mellenbergh y Pieter Vijn publican "The Rasch Model as a Loglinear Model" donde se aprovechan las propiedades de los modelos loglineal para la estimación del modelo de Rasch. Por otra parte la robustez del modelo de Rasch es analizada en un trabajo de Forsyth,

Saisangjan y Gilmer. Estos autores concluyen que los parámetros de los ítems y los parámetros de capacidad permanecen razonablemente invariantes para diferentes tests y grupos de sujetos aún bajo ciertos incumplimientos de las condiciones del modelo.

- 1.981 John P. McIver y Edward G. Carmines publican "Unidimensional Scaling" (en la serie 'Quantitative Applications in the Social Sciences' de Sage). El texto puede considerarse un pequeño manual introductorio a las técnicas de escalamiento unidimensional de Thurstone, Likert, Guttman y Coombs. El mismo año Milton Lodge publica en la misma serie "Magnitude Scaling: Quantitative Measurement of Opinions" que trata especialmente acerca de los métodos de estimación de magnitudes de Stevens para la medición de estímulos sociales. El libro, después de un repaso de las aportaciones de la psicofísica, trata acerca de la aplicación de los métodos de escalamiento al campo de los juicios y las opiniones. Ambos libros tienen su mejor virtud y su principal defecto en su capacidad de explicar claramente los temas más relevantes sin hacer uso apenas de las matemáticas. Sin embargo, pueden resultar muy recomendables para un curso introductorio, y vienen a llenar el déficit de manuales sobre esta temática.
- 1.981 Rand R. Wilcox dedica un trabajo a reconsiderar el uso y la puntuación de tests construidos bajo el procedimiento "Answer-Until-Correct", que ya había sido defendido en 1.950 por Pressey.
- 1.981 Revisión de David J. Weiss y Mark L. Davison bajo el título "Test theory and methods" con tres grandes apartados: teoría clásica de tests, alternativas a la teoría clásica de tests y validez.
- 1.981 Susan S. Schiffman, M.Lance Reynolds, y Forrest W. Young publican su "Introduction to Multidimensional Scaling: Theory, Methods and Applications" (N.Y. Academic Press). El libro considera seis programas disponibles de MDS (MINISSA, POLYCON, KYST, ALSCAL, INDSCAL y MULTISCALE) y las aplicaciones del MDS, sin embargo presenta muy sucintamente el cuerpo teórico del MDS en la primera de sus tres secciones. La segunda, dedicada a una guía práctica de los programas disponibles es la parte más importante del libro. Se presentan dos tablas con los tipos de análisis de los programas comparados que pueden considerarse muy útiles. En conjunto, se trata de una obra que puede considerarse orientada a los usuarios de los programas de escalamiento multidimensional.
- 1.981 Stephen Jay Gould publica "The Mismeasure of Man", una crítica feroz y sesgada escrita en tono polémico, más que bien fundada en investigaciones, contra las tesis hereditaristas y la medición psicológica. La obra es considerada no por su calidad científica si no por que es fiel representante de una crítica ideológica encarnizada a la medición psicológica que toma sus propias hipótesis por verdades probadas y que evita discutir las cuestiones desde el punto de vista de la contrastación empírica rigurosa.
- 1.982 Anne Anastasi publica la quinta edición revisada de "Psychological testing" (McMillan Publishing Co. Inc. N.Y.) La nueva edición incorpora aplicaciones

de la TRI a la construcción de tests, cuestiones relativas a los sesgos de los tests y la aproximación bayesiana a la generalización de la validez. Sin embargo, el capítulo sobre análisis de ítems permanece sin las debidas actualizaciones. Como en ediciones anteriores, lo más específico e interesante del manual de Anastasi son los capítulos dedicados a la descripción de principales tests y baterías de tests. También como en ediciones anteriores no debe buscarse en la obra una exposición matemática de la teoría psicométrica.

- 1.982 Benjamin D. Wrigth y Geoffrey N. Masters, de la Universidad de Chicago, publican un volumen titulado "Rating Scale Analysis" y subtítulo "Rasch Measurement". El libro está expresamente inspirado en la obra de Georg Rasch y contiene algunos capítulos verdaderamente interesantes y didácticos junto a otros donde se ejemplifica el uso de algunos modelos de medida con datos concretos. Después de un capítulo primero, dedicado a tratar los aspectos generales de la medición de un modo atractivo y esclarecedor, el capítulo dos titulado "Examining Data" efectúa una introducción al tratamiento e inspección de los datos desde un punto de vista de medición. Especial interés tienen los capítulos tres y cuatro. El capítulo tres presenta cinco modelos de medida fundados sobre principios del modelo de Rasch. El capítulo cuatro describe cuatro procedimientos diferentes de estimación de los modelos, PROX que puede usarse sin ordenador, PAIR indicado cuando se tienen datos faltantes, CON quizás el más preciso pero que puede generar problemas de cálculo, y UCON que es una aproximación suficientemente precisa a CON. El capítulo cinco trata acerca del, en sus propias palabras, 'control de calidad' en la construcción de una variable introduciendo un índice de "separación" entre los ítems y entre los sujetos escalados. También se ocupan de describir un procedimiento para detectar desajustes, en ítems o en sujetos, al modelo de medida. Los capítulos seis, siete, ocho y nueve son ilustraciones de aplicación de esas técnicas. En conjunto el libro es una pieza importante en el desarrollo de los modelos de medida de Rasch.
- 1.982 Erling B. Andersen dedica un artículo titulado "Latent trait models and ability Parameter Estimation" a la comparación de los métodos de estimación de parámetros de máxima verosimilitud condicional, directo y marginal, fundado en la conexión entre los modelos ITR para ítems dicotómicamente valorados y los modelos para tablas de contingencia.
- 1.982 Fritz Drasgow presenta un artículo titulado "Choice of Test Model for Appropriateness Measurement". El concepto de adecuación de la medida (appropriateness) hace referencia a la detección de aquellos sujetos para los que las puntuaciones en un test no son medidas representativas de su capacidad. Según Drasgow una puntuación puede ser un mal indicador de la capacidad por una variedad de razones que incluyen la escolarización atípica, el que el sujeto haya copiado respuestas o la creatividad excepcional. El trabajo es una aplicación de la teoría de la respuesta al ítem a la cuestión de la adecuación y efectúa comparaciones desde el punto de vista de la adecuación entre el modelo logístico de tres parámetros y el modelo de Rasch. El artículo de Drasgow es en realidad una síntesis de los principales resultados de su tesis doctoral, dirigida por

Michael V. Levine, uno de los pioneros en la elaboración del concepto de adecuación y en su medición.

- 1.982 Isaac I. Bejar y Marilyn S. Wingersky, en un trabajo titulado "A study of Pre-Equating based on Item Response Theory" desarrollan y ponen a prueba la posibilidad de realizar una pre-igualación de dos tests basada en la ITR, con resultados semejantes a los que se obtienen mediante la igualación a posteriori.
- 1.982 Paul W. Holland y D.B. Rubin editan "Test Equating". El libro contiene los trabajos presentados en la "Test Equating Conference" organizada por el Educational Testing Service en abril de 1.980. Los editores presentan una sumaria descripción del campo: "La igualación (Equating) de tests implica a la vez las técnicas para la minuciosa construcción de los tests y los métodos para hacer ajustes estadísticos de las puntuaciones obtenidas de diferentes formas de tests para compensar sus diferencias en dificultad relativa." El libro tiene cuatro partes: análisis de los procedimientos de igualación usados por el ETS, métodos de igualación de la teoría de la respuesta al ítem, nuevos métodos para la igualación de tests y, por último, tópicos relacionados y discusión. Debido a su origen como "actas", el libro no llega a constituir una revisión del campo, como parece ser su propósito declarado, y algunos importantes tópicos como la igualación equipercenil no son tratados específicamente. Sin embargo, buena parte de los procedimientos de igualación son tratados con cierto detalle.
- 1.982 Robert L. Thorndike publica "Applied Psychometrics" (Boston MA. Houghton Mifflin Co.). La obra es un manual que trata tanto de la teoría clásica de tests como de otros tópicos tales como teoría del rasgo latente y teoría de la generalizabilidad, orientado a la aplicación psicométrica. Presenta dos partes muy desiguales, con un adecuado tratamiento de la teoría clásica y un tratamiento de los otros desarrollos poco recomendable. Pueden destacarse los capítulos dos, tres y cuatro que tratan respectivamente de la planificación del desarrollo de un test, de la elaboración de los ítems del tests y de la selección de los ítems. En la parte dedicada al "método de rasgo latente" hay ciertas malinterpretaciones y/o errores conceptuales de cierta magnitud en torno a la interpretación de las curvas características de ítem y sus funciones. La editorial Limusa ha publicado en México en el año 1.989 una versión castellana de este libro que sufre un traductor no psicómetra.
- 1.982 Ronald A. Berk edita su "Handbook of Methods for Detecting Test Bias" un libro que a pesar de ser una recopilación de los trabajos presentados en el "Johns Hopkins University National Symposium of Educational Research" celebrado en 1.980, constituye una ordenada revisión del estado de la cuestión tratando a lo largo de nueve capítulos primero de la definición de "sesgo" (una definición que no llega a hacerse explícita) y después de los diversos métodos de detección de sesgos.
- 1.982 Ronald K. Hambleton edita "Advances in Item Response Theory and Applications" que contiene contribuciones de diversos autores australianos, europeos y americanos entre los que destacan Darrell Bock y F.M. Lord.

- 1.983 Brennan, R.L. publica "Elements of Generalizability theory", una presentación integrada de la teoría de la generalizabilidad.
- 1.983 Charles L. Hullin, Fritz Drasgow y Charles K. Parsons publican "Item Response Theory: Application to Psychological Measurement" (Homewood IL. Dow Jones-Irwin), un libro que pretende ser un manual sobre "teoría de la respuesta al ítem o teoría del rasgo latente" para estudiantes y profesionales de la medición que aprendieron previamente la teoría clásica de los tests. El primer y el último capítulo se ocupan del contexto social de la medición y sus implicaciones. Los capítulos dos y tres son una introducción a la TRI con algunas menciones al desarrollo histórico de la misma. Los capítulos cuatro al ocho se ocupan de aspectos particulares. Debe destacarse el capítulo cuatro que resume la investigación de los autores sobre adecuación (appropriateness) de la medición. Los capítulos cinco y seis se ocupan de los sesgos de los tests y las traducciones de tests, el siete de los tests computarizados (computerized adaptive testing) y el ocho de la cuestión de la dimensionalidad de un test. La información del libro en cuanto a uso técnico se circunscribe al programa LOGIST. Un aspecto interesante de esta obra es que trata de la evaluación de la fidelidad de las versiones traducidas de un test utilizando la TRI.
- 1.983 David J. Weiss edita "New Horizons in Testing: Latente Trait Test Theory and Computerized Adaptive Testing." El libro contiene un capítulo de Levine y Drasgow acerca de la cuestión de la adecuación de las medidas.
- 1.983 Frank B. Baker en "Comparison of Ability Metrics Obtained under Two Latent Trait Theory Procedures" efectúa una comparación de la estimación de parámetros de la TRI bajo los programas de ordenador BICAL y LOGIST.
- 1.983 Fritz Drasgow y Charles K. Parsons en un artículo titulado "Application of Unidimensional Item Response Theory Models to Multidimensional Data" donde ponen a prueba los efectos de violar el supuesto de unidimensionalidad para los modelos de la TRI. La conclusión principal es que los modelos de la teoría de la respuesta al ítem pueden ser aplicados a conjuntos de ítems "moderadamente heterogéneos".
- 1.983 Howard Wainer y Samuel Messick editan bajo el título de "Principals of modern psychological measurement" y el subtítulo de "A festschrift for Frederic M. Lord" un conjunto de trabajos presentados con motivo del setenta cumpleaños de Lord en 1982, en un simposium organizado por el ETS con contribuciones significativas de W.S. Torgerson, A. Anastasi, R.L. Thorndike, N. Cliff, L.R. Tucker, K.G. Jöreskog, F. Samejima, R.D. Bock, J.A. Keats, J.W. Tukey, y M.R. Novick, para citar solo algunos de los autores, colaboradores y discípulos que participaron en el volumen. Además el libro contiene una bibliografía completa de Lord y una extensa nota biográfica que puede contribuir a reconstruir la historia y desarrollo de uno de los principales equipos de trabajo en psicometría.
- 1.983 Isaac I. Bejar publica "Achievement Testing. Recent Advances" en la serie "Quantitative Applications in the Social Sciences" de Sage). Contiene un relativamente extenso capítulo tres sobre "Psychometric Foundations" que considera separadamente el modelo RS (random-sampling) de muestreo

aleatorio de items y la TRI, o modelo de rasgo latente, como la denomina. También se ocupa detalladamente de la cuestión de la administración de tests desde el punto de vista del "adaptive-testing".

- 1.983 Jacob Cohen publica un artículo titulado "The cost of dichotomization" donde pone claramente de manifiesto los perjuicios que provoca la extendida práctica en el contexto de la psicometría de dicotomizar las puntuaciones. "Asumiendo normalidad bivariada con correlación r , dicotomizar una variable por la media provoca una reducción de la varianza explicada de $.647r^2$; y dicotomizar ambas por la media, de $.405r^2$. Esas pérdidas conllevan una reducción de la potencia estadística equivalente a prescindir del 38% y el 60% de los casos bajo condiciones representativas. Si la dicotomización se aleja de la media los costes en la varianza explicada y en la potencia serán todavía mayores." Destacamos el artículo de J. Cohen porque, quizás debido a que una gran parte de la teoría clásica de tests y de los desarrollos en teoría de la respuesta al ítem se han elaborado para items valorados dicotómicamente, ha existido y todavía subsiste una tendencia -que en términos generales puede considerarse nefasta- a dicotomizar las variables en el terreno psicométrico.
- 1.983 Leslie J. Fyans edita "Generalizability Theory: Inferences and Practical Applications" (San Francisco. Jossey-Bass) Este libro se inserta dentro de la serie "New directions for Testing and Measurement" de la editorial Jossey-Bass y se presenta como un intento de exposición sistemática del cuerpo teórico y las aplicaciones de la teoría de la generalizabilidad. El libro, que consta de cinco capítulos, pretende realizar una aproximación al conjunto de la teoría de la generalizabilidad. El capítulo primero "Generalizability Theory: Applications to Program Evaluation" que está escrito por Gillmore contiene, más que un desarrollo de las aplicaciones a la evaluación de programas, una introducción a la teoría de la generalizabilidad. El segundo capítulo, escrito por Cardinet y Allal, es especialmente interesante. Se titula "Estimation of Generalizability Parameters" y ofrece una reformulación de las fórmulas de la teoría de la generalizabilidad que es especialmente útil para la aplicación práctica de la misma. Este capítulo es una valiosa contribución que vuelve más directamente aplicable la teoría. Debe tenerse en cuenta que las aplicaciones de la teoría de la generalizabilidad a problemas prácticos de medición no pueden considerarse muy abundantes, quizás en parte debido al enfoque del libro de Cronbach, Gleser, Nanda y Rajaratnam de 1.972 que no facilitaba los algoritmos de cálculo para diversos diseños de medición. El capítulo tres de esta obra "Multilevel analysis and cross-level inference of validity using generalizability theory" está escrito por el editor de la misma y trata principalmente de la generalización de resultados de investigación a condiciones distintas de las del estudio original. Entre otros aspectos puede destacarse que introduce las ecuaciones de estimación bayesianas para universos de puntuaciones. Webb, Shavelson y Maddahian escriben el capítulo cuatro "Multivariate generalizability theory" que constituye una breve aproximación ejemplificada a esta cuestión. El capítulo cinco "Multivariate generalizability models for tests development from tables of specification, por Jarjoura y Brennan, vuelve sobre el tema del capítulo anterior, pudiendo considerarse una extensión y ampliación del mismo. El libro en su conjunto constituye una valiosa aportación a la teoría

de la generalizabilidad que se suma a la que efectuara Brennan en otro libro del mismo año 1.983.

- 1.983 Mark L. Davison publica su "Multidimensional Scaling" que puede considerarse uno de los intentos más conseguidos por elaborar un libro de texto sobre el campo. El capítulo uno, el dos y el tres pueden considerarse introductorios, destacando la discusión de las medidas de similitud que presenta el capítulo tres. El capítulo cuatro se ocupa del algoritmo métrico de MDS presentado por Torgerson. El capítulo cinco del MDS no métrico. Los capítulos seis y siete de los modelos de diferencias individuales. El octavo se ocupa del uso de las técnicas de componentes principales y regresión en la rotación e interpretación de las dimensiones de un MDS. El capítulo nueve intenta comparar y poner en relación el MDS con el escalamiento de Thurstone, y con el cluster y el análisis factorial.
- 1.983 Peter Dunn-Rankin publica "Scaling Methods", un libro que está orientado como un manual sobre escalamiento para estudiantes e investigadores con enfoque práctico más que teórico. El volumen, después de unos capítulos introductorios sobre medición, se reparte en tres partes principales. La primera se dedica a los procedimientos del escalamiento unidimensional, la segunda al "clustering" y la tercera al escalamiento multidimensional. Una última parte presenta una útil y extensa guía de los programas disponibles en el campo del escalamiento multidimensional.
- 1.983 Ronald K. Hambleton edita "Applications of Item Response Theory", un libro que recopila trece trabajos de diversos autores sobre la TRI, ocho de los cuales se dedican propiamente a aplicaciones, mientras que los cinco primeros actúan como una presentación general. Los trabajos de Swaminathan tratan sobre la estimación de los parámetros de la TRI. El de Traub enfatiza la necesidad de poner a prueba diversos modelos de funciones TRI sugiriendo que en buena parte de los casos la solución de Rasch no es la más adecuada. Wigersky presenta el uso del programa LOGIST. Hambleton y Murray presentan en su capítulo un conjunto de métodos para evaluar la bondad de ajuste de TRI a un conjunto de datos. En conjunto el libro presenta un número de aplicaciones de la TRI con la omisión del 'computerized adaptive testing'. Una conclusión general a lo largo de sucesivos capítulos es la superioridad del modelo logístico de tres parámetros sobre el modelo de Rasch en su ajuste a los datos. También en 1.983 Ronald K. Hambleton publica un artículo donde extiende el uso de los modelos de la Teoría de la Respuesta al Ítem a la evaluación referida al criterio.
- 1.983 Steven E. Rigdon y Robert K. Tsutakawa publican "Parameter estimation in latent trait models". En este trabajo se estudia una aproximación a la estimación de parámetros bajo la teoría del algoritmo EM con aplicación al modelo de Rasch.
- 1.983 Steven J. Osterlind publica "Test Item Bias". Un libro que constituye "una discusión acerca de los procedimientos estadísticos apropiados para explorar las propiedades psicométricas de sesgo en ítems de tests estandarizados." Se define un ítem insesgado como aquel para el que la probabilidad de éxito es la misma para sujetos igualmente capaces de la

misma población, independientemente de su pertenecía a subgrupos de características particulares. El libro trata los cinco procedimientos de detección de ítems sesgados más ampliamente usados: el análisis de la varianza, las dificultades de los ítems transformados, chi cuadrado, las curvas características de ítem y las respuestas distractoras, dedicando un capítulo a cada una de las técnicas después de un apreciable capítulo introductorio. El libro presenta estos temas con un alto nivel de claridad y estructuración, proveiendo las herramientas conceptuales suficientes para una aplicación de los conceptos a casos reales.

- 1.984 Dato N.M. de Gruijter y Leo J. Th. van der Kamp publican "Statistical Models in Psychological and Educational Testing" (Lisse, Swetts y Zeitlinger). La obra es un intento de integrar en un libro la teoría de los tests desde la teoría clásica hasta la TRI y sus aplicaciones. El capítulo uno trata de la fundamentación de la medición, ocupándose de la cuestión de los estadísticos admisibles. El capítulo dos trata de los modelos fuerte y débil de la puntuaciones verdaderas. El capítulo tres se dedica a las técnicas de estimación de parámetros. El cuatro trata la teoría clásica de tests. El cinco se ocupa de la teoría de la generalizabilidad. El seis de los modelos binomiales. El siete hace referencia a los análisis de estructuras de covarianza, poniendo especial atención en el LISREL. El capítulo ocho se ocupa de la validez. Los capítulos nueve y diez se dedican a la teoría de la respuesta al ítem. Por último los capítulos once, doce y trece se ocupan de las aplicaciones de la teoría de la respuesta al ítem.
- 1.984 David H. Barlow y Michel Hersen publican "Single Case Experimental Designs" obra clave en el marco del diseño para el uso de medición en la investigación con N=1. La obra será traducida al castellano en 1.988.
- 1.984 Revisión de Forrest W. Young sobre "escalamiento" con tres grandes apartados: teoría de datos, escalamiento de datos de disimilaridad por modelos de distancia, y escalamiento de datos multivariados por modelos lineales y bilineales.
- 1.984 Ronald A. Berk edita "A Guide to Criterion-Referenced Test Construction" (Baltimore. The Johns Hopkins University Press). Este libro es una versión actualizada y aumentada del que el mismo Berk editara en 1.980 con trabajos presentados en el 1.978. Los seis autores que participaron en la obra original tienen también aquí capítulos, a los que se añaden nuevas contribuciones hasta un total de once capítulos. Puede destacarse el capítulo uno escrito por Anthony J. Nitko y titulado "Defining 'Criterion-Referenced Test'", junto a los de Popham sobre descripción de dominios, Roid sobre generación de ítems, Millman sobre tests individualizados, Shepard sobre estandars y Hambleton sobre la validez.
- 1.985 Revisión de Ross E. Traub e Y. Raymond Lam sobre "Latent Structure and Item Sampling Models for Testing" donde se presenta un marco ordenador del campo del análisis de la estructura latente incluyendo dos grandes familias, la de la teoría de la respuesta al ítem y la del análisis de clase latente. Quizás lo más importante de esta excelente revisión es que provee de un armazón conceptual en el que pueden ordenarse adecuadamente los conocimientos y trabajos acerca del análisis de la estructura latente.

Resulta interesante el modo en que los revisores utilizan estas etiquetas distinguiendo la TRI como una gran rama del análisis de la estructura latente, en lugar de tomarlos por sinónimos mostrando preferencia por la denominación de TRI, como ha sido frecuente en una serie de trabajos en tiempos recientes.

- 1.985 Hambleton, R.K. y Swaminathan, H. publican "Item response Theory: Principles and applications" (Boston MA, Kluwer). El libro persigue, a decir de los autores, tres propósitos, presentar la TRI de un modo accesible no técnico, ofrecer un tratamiento comprehensivo de los modelos y sus aplicaciones, y expresar los pasos específicos necesarios para usar los modelos en determinadas aplicaciones. El primer capítulo hace una breve presentación histórica de la ITR y una presentación sumaria de la misma a partir de las dificultades de la teoría clásica. El capítulo dos se dedica a los supuestos de la TRI, y el tercero trata acerca de los modelos de cci. Los capítulos siguientes tratan diversos aspectos de la teoría de la respuesta al ítem, incluyendo las funciones de información (Cap. 6), los métodos de estimación de los parámetros de los ítems (Cap. 7), el ajuste de las cci a los datos (Cap. 8), el ajuste del modelo a los datos (Cap. 9), la igualación de tests bajo la TRI (Cap. 10) y algunas aplicaciones de la TRI (Cap. 13).
- 1.985 Frank B. Baker publica "The basics of Item Response Theory" que es, a decir de la crítica, uno de los mejores ensayos para poner la TRI a nivel introductorio de modo que pueda ser introducida adecuadamente en la docencia. Los capítulos uno y dos introducen la noción de cci. El capítulo tres expone un método de estimación de los parámetros. El capítulo cuatro introduce el concepto de c.c.t. (curva característica del test). El capítulo cinco ofrece un método para la estimación de la capacidad de un único sujeto examinado. El capítulo seis se ocupa de la función de información. El siete trata de la calibración de tests bajo la TRI. El capítulo ocho muestra como utilizar un banco de ítems para elaborar tests con diferentes propósitos. Como innovación el libro va acompañado de un disco para Apple II+ que, para ese restringido grupo de usuarios, complementa de modo valioso el aprendizaje evitando los cálculos tediosos y permitiendo ensayar ejemplos.
- 1.985 Nueva edición de los "Standards for educational and psychological testing" de la American Psychological Association. Entre las novedades destaca la atención que se presta al 'equating', ausente en la edición anterior de 1.974.
- 1.986 Charles Lewis publica en Psychometrika una revisión de la teoría de tests en esa revista durante los veinticinco años pasados.
- 1.986 Joel Michel publica un trabajo titulado "Measurement Scales and Statistics: A clash of paradigms" donde se enfrenta con la polémica sobre la relación entre escalas y estadísticos permisibles que iniciara Stevens, y que recientemente habían reabierto los trabajos de Gaito de 1.980 "Measurement scales and statistics: resurgence of an old misconception" y de Townsend y Ashby de 1.984 "Measurement scales and statistics: The misconception misconceived". Michell identifica tres grandes paradigmas de teorías de la medición, el representacional, el operacional y el clásico, y postula que la controversia es fruto del choque de estos diferentes

paradigmas de la medición. Según Michell la teoría representacional si que implicaría una relación entre escalas y estadísticos, restringiendo el tipo de estadísticos permisibles, aunque no necesariamente según los postulados de Stevens. Sin embargo, las teorías operacional y clásica no supondrían este tipo de relación.

- 1.986 P. Kline publica "A Handbook of Tests Construction" (London. Methuen.) Se trata de un manual "práctico" de nivel elemental que introduce en los aspectos básicos del tema.
- 1.986 Anne Anastasi escribe la revisión "Evolving concepts of test validation" (Ann. Rev. Psychol.)
- 1.987 Allan L. McCutcheon publica "Latent Class Analysis". El análisis de clase latente se ocupa de tratar con variables latentes inobservadas concebidas como categóricas, siendo esta la principal diferencia con la teoría de la respuesta al ítem. El análisis de clase latente permite identificar variables latentes discretas a partir de variables observadas discretas. Los capítulos del libro se ocupan sucesivamente de efectuar una introducción a la lógica de la variables latentes, explicar el modelo y el supuesto de independencia local, la estimación de las variables latentes categóricas, el análisis de los patrones de respuesta y la comparación de estructuras latentes entre grupos.
- 1.987 Reimpresión de la obra de Harold Gulliksen "Theory of Mental Tests".
- 1.987 Drasgow y Guertler presentan varios índices de adecuación derivados de la TRI.
- 1.987 Lee J. Cronbach interviene en la polémica cuestión de la evaluación de la hipótesis moduladora con un artículo titulado "Statistical Test for Moderator Variables: Flaws in Analysis Recently Proposed" en el que defiende el procedimiento tradicional de regresión múltiple modulada convencional y rechaza los procedimientos elaborados por Morris, Sherman y Mansfield (1.986), y por Bobko (1.986). La crítica de Cronbach no es la única que ha recibido el trabajo de Morris, Sherman y Mansfield. En este mismo año y en el mismo número del Psychological Bulletin del artículo de Cronbach, otro trabajo, de Dunlap y Kemery pone de manifiesto que el resultado de que la regresión múltiple modulada por mínimos cuadrado ordinarios infraestima los efectos moduladores es incorrecta para los propios datos de Morris y colaboradores, añadiendo que los efectos detectados al utilizar regresión por componentes principales deben ser un artefacto de este tipo de regresión.
- 1.988 David Andrich publica "Rasch Models for Measurement" en la serie "Quantitative Applications in the Social Sciences" de Sage. A diferencia de lo que es común en los libros de esta utilísima serie, el volumen de Andrich no puede constituir un trabajo introductorio en el tema a pesar, y quizás precisamente por ello, del gran esfuerzo del autor por sintetizar con gran brevedad los principales conceptos. En contra de lo que sugiere su título el volumen no se circunscribe exactamente a los modelos de Rasch contando con dos capítulos introductorios muy generales, una serie de desarrollos

más allá del modelo de un parámetro, y una interesante consideración de las cuestiones de estimación y tests de ajuste.

- 1.988 En Noviembre ve la luz la versión siete del programa LISREL de Jöreskog y Sörbom que permite estimar, con mejoras sobre versiones anteriores, modelos de medida exploratorios y confirmatorios y modelos de ecuaciones estructurales simultáneamente tomados.
- 1.988 George A. Gescheider publica una revisión sobre escalamiento psicofísico que ordena los contenidos del campo bajo los siguientes grandes apartados: componentes del experimento de escalación, estimación de magnitudes, estimación de magnitudes absoluta, sesgos de respuesta y relación entre diferentes escalas.
- 1.988 Mark L. Davison y Anu R. Sharma en un trabajo titulado "Parametric Statistics and Levels of Measurement" vuelve sobre la cuestión de la relación entre escalas y estadísticos permisibles. En esta polémica que iniciara Stevens en 1.951 las dos posturas enfrentadas sostienen una distinta relación entre estadísticos y escalas. Townsend y Ashby (1.984) por ejemplo, defienden que los estadísticos paramétricos son inadecuados para datos ordinales. En la postura opuesta Borgatta y Bohrnstedt (1.980), Gaito (1.980) y Terwilliger, en un trabajo inédito de 1.980, sostienen que el nivel de medida resulta irrelevante. Bajo la consideración de este contexto, y de las aportaciones de Michell (1.986) el trabajo de Davison y Sharma da una respuesta matizada que abreviadamente puede resumirse en el sentido de que, cumpliendo determinados supuestos, los estadísticos paramétricos pueden ser indicados en el nivel de medida ordinal.
- 1.988 Phipps Arabie, J. Douglas Carroll y Wayne S. DeSarbo publican "Three-Way Scaling and Clustering" El libro discute en detalle el método INDSCAL (Individual Differences Scaling) de Carroll y Chang (1.970) y el método INDCLUS (Individual Differences Clustering) de Carroll y Arabie (1.983), además de otros métodos menos divulgados de escalamiento multidimensional, y puede considerarse una extensión más avanzada del volumen introductorio de Kruskal y Wish "Multidimensional Scaling" de 1.978.
- 1.988 Z. Govindarajulo publica un trabajo titulado "Alternative methods for combining several tests scores" donde revisa los diversos procedimientos para la combinación de las puntuaciones de tests, sugiere otros nuevos y efectúa una comparación entre los mismos.