

## **PERSPECTIVA HISTORICA DE LA TEORIA DE LA RESPUESTA AL ITEM.**

JOSÉ LUIS MELIA,  
JOSÉ MANUEL TOMAS  
AMPARO OLIVER.  
Universitat de València

### **RESUMEN**

El artículo revisa los desarrollos de la teoría de la Respuesta al Item (TRI) desde una perspectiva histórica. Se organiza en siete puntos. El primero, sitúa los orígenes históricos de la teoría y el lugar que ocupa frente a la teoría clásica de tests. El segundo da cuenta de los desarrollos conceptuales producidos principalmente en la década de los cincuenta. Durante la década de los sesenta, tercer punto, se resalta la generación de distintos modelos todavía vigentes en la investigación. El cuarto punto se corresponde con la década de los sesenta, cuya principal característica podría ser la aparición de programas estadísticos preparados para implementar la TRI en situaciones reales y su utilización en aplicaciones. La década de los ochenta, quinto punto, se caracteriza por una reintegración teórica, aparición de manuales introductorios y una profundización en las aplicaciones. En el punto seis se revisan las aportaciones realizadas a la TRI desde nuestro país. Cierra el artículo un apartado de conclusiones.

### **ABSTRACT**

Present article reviews in seven points Item Response Theory (IRT) from an historic perspective. First point, introduction, gives historical origins of IRT and its position in front of Classical Test Theory inside Psychometry. Second point presents the conceptual basis of item response theory that appeared in fiftieths. Third point presents the models widely used in IRT that were developed basically during sixtieths. Seventieths, which forms fourth point, have a defintory characteristic, the presentation of statistical software for the IRT implementation. Eightieths -fifth point- represent a theoretic reintegration and a prolusion of aplicaciones. In point six the IRT in Spain is presented. The article is closed with a point devoted to conclusions and state of the art.

La teoría clásica de tests, generada a partir de los trabajos sobre la descomposición de la puntuación empírica en puntuación verdadera más error de medida de Spearman, ha dominado durante los primeros cincuenta años del siglo el panorama psicométrico. Sin embargo, otro cuerpo teórico diferente ha ido apareciendo gradualmente para competir, ayudar o sustituir, según el caso, a la teoría clásica. Esta moderna teoría de la medición

psicológica y educacional se ha denominado de diversos modos a lo largo de los treinta o cuarenta años de desarrollo: teoría del rasgo latente, teoría de la curva característica o el acuñado más recientemente y ampliamente aceptado de teoría de la respuesta al ítem.

Frente a los pocos supuestos que realizaba la teoría clásica de tests, la teoría de la respuesta al ítem (TRI) propone un buen número de fuertes supuestos matemáticos que debieran encontrarse en los datos para poder implementar la teoría. Si estos supuestos se cumplen, las ventajas de la TRI frente a la teoría clásica son claras y podemos resumirlos con Lord y Stocking (1989) en tres puntos: a) es posible caracterizar un ítem independientemente de la muestra que pase el test o el ítem, b) es posible caracterizar a una persona en el atributo o atributos medidos por el test independientemente de la muestra de ítems que se le administre, y c) es posible predecir propiedades de un test antes de su administración. En una frase, la TRI permite a diferencia de la teoría clásica mediciones de los sujetos independientes de los ítems concretos administrados y, además, mediciones de los ítems independientes de la muestra de personas que los contestan. De este modo la TRI vino a satisfacer la vieja aspiración métrica de obtener mediciones no relativas al grupo, independientes de la muestra e independientes del instrumento concreto.

Pese a las evidentes ventajas que tiene este tipo de medición invariante, lo cierto es que hasta fechas bien recientes pocas han sido las aplicaciones de estos desarrollos psicométricos. La mayor parte de éstas se circunscriben a tests de muy largo alcance que aluden básicamente a medición de aptitudes y capacidades, y desde luego generalmente se han venido desarrollando en EEUU y fundamentalmente desde agencias oficiales dedicadas a la creación de tests de ámbito estatal como el Educational Testing Service.

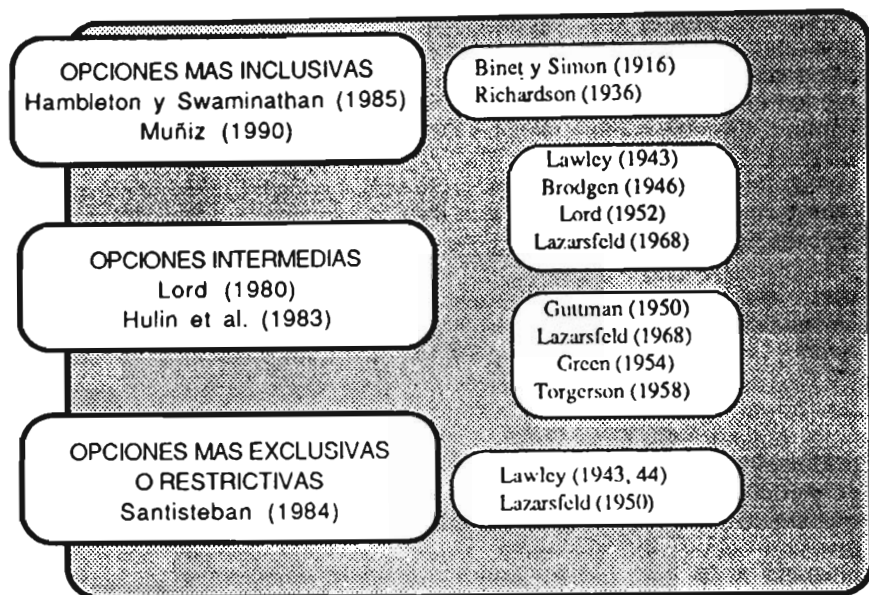
El objetivo de este trabajo es hacer un breve repaso con cierta perspectiva histórica de lo que entendemos como aportaciones más relevantes tanto teóricas como aplicadas a la TRI. La mayor parte de citas que recogeremos, como suele ser habitual son de literatura anglosajona, y más concretamente estadounidense. Al final, se recogerán algunas de las aportaciones realizadas desde nuestro país a estos modernos desarrollos de la teoría de tests.

## 1. INTRODUCCION

Algunos autores afirman que la TRI surge a lo largo de los años cincuenta y sesenta como reacción a la teoría clásica de tests, de algún modo justificada por las deficiencias de esta última que sí se podían resolver desde la nueva teoría (Van der Linden y Hambleton, 1982; Van der Linden,

1986). Esto es, ven a la Teoría de Respuesta al Ítem de algún modo como una alternativa a la teoría clásica. A diferencia de esta postura, uno de los principales teóricos de la TRI, F. M. Lord (1980) afirma al respecto del cuerpo teórico de la TRI y su relación con la teoría clásica de tests: "nada...va a contradecir ni los supuestos ni las conclusiones básicas de la teoría clásica de tests. Se harán supuestos adicionales, estos nos harán posible contestar cuestiones que la teoría clásica no puede responder... completamos más bien que contradecimos la teoría clásica de tests" (Lord, 1980). Otros autores toman una postura intermedia entre estos dos planteamientos. Hulin, Drasgow y Parsons (1983) dicen al respecto que les parece adecuado ver a ambas teorías como parcialmente solapadas, y afirman de modo contundente: "los modelos de la TRI hacen supuestos más fuertes que los hechos en la teoría clásica y por tanto obtienen resultados más fuertes" (Hulin, Drasgow y Parsons, 1983).

Los orígenes de la TRI pueden fecharse en 1943 cuando se publica el trabajo titulado "On problems connected with items selection and test construction" de D. N. Lawley. Se presenta una primera descripción del modelo de ojiva normal, base de la TRI, como función que pone en relación un rasgo latente tomado como continuo con la probabilidad de respuesta correcta a un ítem (Meliá, 1990). Este trabajo aparece en numerosas publicaciones como el punto de partida de los Modelos de Rasgo Latente. Muchos autores identifican la obra de Lawley (1943), de Tucker (1946), de Lord (1952, 53) y Birnbaum (1957,58) como el inicio de la TRI. Otros autores (Hambleton y Swaminathan, 1985; Muñoz, 1990) optan por una opción más inclusiva que situaría como primer antecedente de la TRI a Binet y Simon (1916) y a Richardson (1936). Lord sitúa a Brodgen (1946), Lawley (1943), Lord (1952) y Lazarsfeld (1968) como precursores de la teoría. Otros autores (Hulin et al, 1983) sitúan los inicios indirectos de la actual TRI en Guttman (1950), Lazarsfeld (1950), Green (1954) y Torgerson (1958). En una forma más exclusiva, que sitúa como antecedentes tan solo a las formulaciones que de algún modo permanecen vigentes en la actualidad, Santisteban (1984) identifica a Lawley (1943, 44) como precursor y sitúa a Lazarsfeld (1950) como el primero en realizar un estudio formal y también el primero en introducir el término rasgo latente. Un esquema de estas tres posturas puede consultarse en la figura 1.



*Figura 1. Orígenes de los modelos de rasgo latente, según autores.*

## 2. DÉCADA DE LOS CINCUENTA, CONFIGURACION TEORICA.

En 1950 aparece el libro de Stouffer, Guttman, Suchman, Lazarsfeld, y Star "Measurement and Prediction". Este es un libro de importancia para la psicometría por diversos motivos, si bien para el tema de este trabajo es sobre todo relevante la aportación de Lazarsfeld (capítulos 10 y 11) uno de los primeros trabajos sobre teoría de la estructura latente. Lazarsfeld aporta la denominación que se impondrá durante los treinta años siguientes, introduce el término "Rasgo Latente" y el concepto de curva característica del ítem. El concepto de curva característica o función de respuesta al ítem es el concepto central de la teoría, alrededor del cual giran todos los desarrollos. Esta curva relaciona la probabilidad de una respuesta a un ítem, habitualmente dicotómico, condicionada al nivel del rasgo latente que posee el sujeto. En este mismo libro, como contribución de Guttman encontramos el modelo de escalamiento perfecto el cual presenta una cuantificación distinta a la anterior de esta misma relación rasgo-probabilidad de respuesta (Hulin; Drasgow y Parsons, 1983). El modelo de Guttman es un modelo determinístico, sin embargo, los desarrollos posteriores de la TRI se

orientarían hacia modelos estocásticos, precisamente presentados por Lazarsfeld en ese trabajo.

En 1951 Frederic Mather Lord, quien es con seguridad el mayor impulsor de la Teoría de Respuesta al Ítem, lee su tesis "A theory of tests scores" en la Universidad de Princeton. Esta tesis, la cual es fruto del trabajo realizado por Lord en TRI desde el año 1941, junto con el libro "Theory of Mental Tests Scores" de Gulliksen serán los dos pilares teóricos en los que se apoyará el Educational Testing Service en los siguientes años. Estos dos trabajos se enmarcarían en TCT y en TRI, respectivamente. La tesis doctoral de Lord se publicará en Psychometrika como la monografía nº 7 (1952), recogiendo los más importantes aspectos teóricos de la TRI formulados hasta entonces. A partir de esta aportación de Lord, queda definido el concepto presentado por Lazarsfeld de curva característica mediante la utilización de la distribución normal acumulada o función de ojiva normal.

La utilización de la función de ojiva normal se hace común hasta que Birnbaum (1957, 1958) presenta una nueva función matemática para describir la función característica del ítem, la función logística, derivada desde el campo del bioensayo (Muñiz, 1990). Esta función logística, que difiere levemente a efectos prácticos de la ojiva normal, presentaba sin embargo, una mayor tratabilidad matemática por lo que tendió a sustituir a la de ojiva normal en la investigación. La función logística, al igual que la de ojiva normal, queda determinada por tres parámetros: la dificultad del ítem, la capacidad de discriminación del ítem, y el parámetro de conjetura que da cuenta de la probabilidad de responder por azar. El modelo determinado por estos tres parámetros será conocido como modelo logístico o de ojiva de tres parámetros, y constricciones en los valores de los parámetros darán nuevos modelos de dos y un parámetro (por ejemplo el modelo de dos parámetros asigna valor cero al parámetro de conjetura).

Las funciones logísticas y de ojiva no son los únicos que en esos años relacionan el rasgo latente con la probabilidad de respuesta al ítem. Un nuevo modelo es formulado por Lazarsfeld en 1959, buscando una relación lineal. En la práctica, la existencia de probabilidades negativas y mayores de uno hacen inconveniente este tipo de modelos para los propósitos de la TRI (Hulin, Drasgow y Parsons, 1983).

En cuanto a trabajos de índole más aplicada contamos en estas fechas con trabajos en torno a la estimación de los parámetros. En un memorándum de investigación del Educational Testing Service de 1951 Tucker introduce un método de máxima verosimilitud para la estimación de parámetros desde el modelo de ojiva normal. Por otro lado, Lord (1953) introduce otro método de máxima verosimilitud diferente al de Tucker para la estimación de parámetros también para el modelo de ojiva.

### 3. DÉCADA DE LOS SESENTA, DESARROLLO DE MODELOS BÁSICOS EN TRI.

Tras los inicios en los cuarenta o cincuenta, la década de los sesenta supone una expansión teórica importante, con la generación de nuevos modelos y con la aparición de soluciones teóricas importantes.

Independientemente de los desarrollos que producía Lord en torno a los modelos de ojiva o Birnbaum con los logísticos en EEUU, en 1960 el matemático danés Rasch escribe el libro "Probabilistic models for some intelligence and attainment tests" donde presenta un modelo de un sólo parámetro. Este modelo, que lleva su nombre, quizá haya generado la mayor cantidad de trabajos teóricos y prácticos en el campo. Lo caracteriza que la respuesta del sujeto sólo se ve influida por dos parámetros, uno del ítem que se le presenta y el otro de la capacidad del sujeto. Tras la aparición del modelo logístico de tres parámetros este modelo logístico de Rasch queda absorbido como un caso especial del último en el que se asume no hay probabilidad de contestar al azar (conjetura igual a cero) y todos los ítems discriminan lo mismo (parámetro de discriminación igual a constante). La viabilidad de un modelo que realiza asunciones tan fuertes en la forma de las curvas características (constricciones en los parámetros) todavía sigue generando fuertes polémicas en la literatura. Así, en torno a B. Wright en la Universidad de Chicago y a Anderson y Masters en Europa se ha defendido la utilidad práctica y la adecuación a los datos del modelo de Rasch y sus derivados. Mientras tanto, Lord y la mayor parte de sus discípulos (por ejemplo Samejima) afirman la superioridad del modelo de tres parámetros como más adecuado en casi todas las condiciones de pase de tests.

Pronto el modelo de Rasch va a ponerse en aplicación debido fundamentalmente a su facilidad de manejo, por los pocos parámetros a estimar, y por la existencia de estadísticos suficientes. En este sentido, B. Wright y N. Panchapakesan (1969) publican el artículo "A procedure for sample free item analysis". Este artículo vuelve a poner de manifiesto que las estimaciones de los parámetros del ítem y el nivel en el rasgo latente (capacidad) de la persona en el modelo de Rasch se producen independientemente de la muestra y de los ítems concretos del test administrado, como ya había sido adelantado por Rasch (1960). Esto es, se consigue fácilmente la deseada invarianza de los parámetros del ítem y de la capacidad de la persona. Además, los autores presentan un estadístico ji-cuadrado útil para comprobar la bondad de ajuste del modelo a los datos.

Al margen de los modelos de Rasch prosiguen los desarrollos teóricos sobre los modelos logísticos promovidos por Birnbaum y Lord. En 1961 Birnbaum en la revista *Annals of Mathematics* prueba matemáticamente las propiedades de eficiencia de los estimadores máximo verosímiles que serán

los más utilizados hasta la actualidad en el proceso de estimación desde la TRI.

Por otro lado, los modelos logísticos crecen al presentarse nuevas formulaciones. McDonald (1967) propone un modelo logístico de cuatro parámetros. Este modelo, de interés puramente teórico (Barton y Lord, 1981) pretende explicar los casos en que los individuos de alta capacidad no responden correctamente a ítems fáciles. Este parámetro representaría la probabilidad (distinta de uno) de contestar correctamente el ítem para sujetos de capacidad muy alta. Además Fumiko Samejima, discípula de Lord en el Educational Testing Service, genera una gran variedad de submodelos de la TRI. Entre estos, el introducido en 1969 (Modelo de respuesta clasificada o graduada) o el presentado en 1972, que coincidiría en el tiempo con el elaborado por Bock (Modelo de Respuesta Nominal). Los modelos presentados por Samejima se basan en la lógica de que se puede extraer información valiosa también en las respuestas incorrectas a los ítems y permiten considerar cada opción de respuesta al ítem, caso de haberlas, de modo separado. En 1973 esta misma autora presentó un modelo adecuado cuando la escala de respuesta era continua que resulta ser un caso límite del modelo de respuesta graduada.

En 1968 se producen dos aportaciones teóricas valiosas en forma de libro. Por un lado, Lazarsfeld y Henry publican el libro "Latent Structure Analysis". Se describen los distintos modelos de estructura latente elaborados hasta el momento, describiéndose los puntos en común y diferencias entre los modelos de clase latente y el escalamiento de Guttman, entre otros.

De mucho más alcance que el primero, Lord y Novick (1968) producen "Statistical Theories of Mental Tests Scores". Este libro ha sido denominado el "suceso señal de la década" (Bock y Wood, 1971), el evento de mayor impacto en el campo de la teoría de los tests (Lewis, 1986), "un gran paso para llevar a la Teoría Clásica y la Moderna hacia un mejor balance" (Bock y Wood, 1971), o con menor entusiasmo "un gran mausoleo" (Lumsden, 1976). Sin duda éste es el desarrollo base para los modelos de TRI, a excepción del Modelo de Rasch. Pero el libro no se circunscribe al ámbito de la TRI, sino que abarca también la teoría clásica de tests y la teoría fuerte de las puntuaciones verdaderas. A propósito de la TRI, son de destacar los capítulos firmados por Allan Birnbaum en los que se analizan, entre otros tópicos, el modelo logístico, la estimación de parámetros, las escalas para medir el rasgo latente, las funciones de información, y en resumidas cuentas todos los desarrollos teóricos de la TRI que aún hoy permanecen vigentes.

#### 4. DÉCADA DE LOS SETENTA, INFORMATICA Y ESTUDIOS DE APLICACION.

La revisión de Bock y Wood publicada en el *Annual Review of Psychology* en 1971 es reflejo de la influencia del libro de Lord y Novick (1968) donde se culmina el proceso de consolidación teórica de los modelos bases en TRI iniciado en la década de los cincuenta, y sobre todo durante los sesenta. Los trabajos más relevantes de la década de los setenta tienen una índole más aplicada, centrada sobre todo en problemas concretos de implementación de los modelos, estimación y ajuste a los datos.

Este nuevo interés por la implementación aparece en gran parte debido a que empiezan a aparecer programas informáticos, sin los cuales era prácticamente imposible la aplicación práctica real. Quizá, el más influyente programa estadístico haya sido el LOGIST, firmado por Wood y Lord (1976) surgido en el seno del Educational Testing Service. Aunque estima con procedimientos de máxima verosimilitud conjunta la capacidad y parámetros del ítem para el modelo de tres parámetros, es capaz de calcular también para modelos de uno y dos parámetros con restricciones. Además es capaz de funcionar aún con respuestas omitidas e ítems no intentados (Hambleton y Swaminathan, 1985). Otros programas de ordenador que aparecen en esta década son el DATAGEN (Hambleton y Rovinelli, 1973), BICAL (Wright y Mead, 1976), y ANCILLES (Urry, 1976) (Santisteban, 1990). El programa BICAL analiza el modelo de un sólo parámetro y provee tests de ajuste para los parámetros del ítem y de la persona. ANCILLES Y OGIVA realizan estimación de máxima verosimilitud conjunta para el modelo de tres parámetros (Hambleton y Swaminathan, 1985).

Con los programas ya disponibles empiezan a producirse implementaciones en grandes muestras y estudios de comparación entre programas. Ejemplo de este último tipo de estudios es el artículo que M. J. Ree publica en la revista *Applied Psychological Measurement* que compara cuatro programas para la estimación de los parámetros de la función de respuesta al ítem usando el modelo de tres parámetros. Su conclusión es que ninguno de los cuatro ANCILLES, LOGIST, OGIVIA, y Transformations parecía consistentemente mejor que los restantes.

La estimación de los parámetros reúne gran número de investigaciones como las de Andersen y Madsen (1977) y Andersen (1977) sobre estadísticos suficientes en el modelo de Rasch. También en la estimación desde el modelo de Rasch, Kearns y Meredith (1975) presentan un acercamiento desde la estimación Bayesiana y Cohen (1979) un procedimiento no-iterativo para calcular la capacidad y la dificultad Dentro de



los modelos logísticos de dos y tres parámetros Samejima (1977) presenta un tipo de estimación máximo verosímil cuando la capacidad de los sujetos ha sido medida previamente, mientras Jensema (1976) aborda la estimación desde los índices clásicos y su transformación a parámetros de la TRI, y Bejar y Weiss (1979) plantean nuevos métodos de puntuación

En cuanto al ajuste datos-modelo Andersen (1973) presenta un test para comprobar la bondad de ajuste del modelo de Rasch. Otros artículos en torno a este tema son el de Keats (1974) y Dinero y Haertl (1977). Keats propone un modelo definido por la dificultad y la conjetura para evitar el problema del mal ajuste que presenta el modelo de Rasch cuando existe adivinación (probabilidad de contestar al azar). Y Dinero y Haertl investigan la aplicabilidad del modelo de Rasch cuando los ítems de un test incumplen el supuesto de que sus discriminaciones sean iguales. Las conclusiones van en la línea de la robustez del modelo frente a moderadas violaciones del supuesto.

En aplicaciones más concretas de la TRI, la igualación vertical generó polémica entre Slinde y Linn, por un lado, y Gustafson, por otro, en el *Journal of Educational Measurement*. Slinde y Linn (1977) firman un artículo en el *Journal of Educational Measurement* que trata sobre la adecuación del Modelo de Rasch para la igualación vertical. Los datos de Slinde y Linn sugieren la existencia de algunos problemas en el uso del modelo de Rasch para la igualación vertical. Exponen estos autores que quizá sean necesarios más parámetros del ítem para esta tarea. Estos resultados de Slinde y Linn provocaron una discusión por parte de Gustafson quien en 1979 y en la misma revista publica el artículo "The Rasch Model in vertical equating of tests: A critique of Slinde and Linn". En este artículo explica que quizá las opiniones de los dos autores se basan en un diseño de investigación con problemas, aún cuando reconoce los problemas inherentes a la igualación vertical mediante el Modelo de Rasch. Tras esta crítica de Gustafson, Slinde y Linn (1979, "A note on vertical equating...and tests of quite different difficulty") repiten su experimento pero con otro conjunto de datos y encuentran los mismos resultados hallados en su primera experiencia. Estos no han sido los únicos trabajos sobre igualación vertical mediante el Modelo de Rasch (Rentz y Bashaw, 1977)

Otro tópico básico dentro de las aplicaciones es el de tests a la medida. Un estudio representativo sobre los tests adaptativos o a la medida puede ser el de Samejima, donde se trata el uso de las funciones de información en este tipo de aplicación. El artículo de Samejima no sólo es importante por su aplicabilidad a problemas concretos que aparecen en el diseño de test a la medida, sino porque presenta una formulación matemática nueva de la función de información del ítem que sería general para los modelos binarios y politómicos de la TRI. La definición anterior de

Birnbaum de función de información sería un caso particular de la presentada por la autora.

Por último, a finales de los setenta sigue abierto el debate sobre la ubicación de la TRI entre la teoría de tests y las técnicas de escalamiento (Andrich, 1978; Brodgen, 1977; Lumsden, 1980; Weiss y Davison, 1981).

## 5. DÉCADA DE LOS OCHENTA, REINTREGACION TEORICA Y APLICACIONES.

Dos son los aspectos que pueden caracterizar la década de los ochenta. Por un lado, se produce una gran aparición de libros de carácter introductorio y de reintegración de los desarrollos teóricos, en un intento de hacer más accesible la teoría. Por otro, se siguen dedicando grandes esfuerzos a aplicaciones de la TRI, especialmente en campos en los que parece que sus resultados son superiores a los obtenidos mediante la teoría clásica. En este sentido, los tests a la medida, la medición referida al criterio, igualación vertical y horizontal, bancos de items, y estudio de sesgos entre otras aplicaciones han merecido atención. En torno a estos dos ejes revisaremos algunas de las contribuciones más relevantes

En esta última década nos hemos inclinado a revisar básicamente libros, frente al mayor número de artículos de las décadas anteriores. Esto debido a varios motivos, por un lado aparecen mayor número de ellos, algunos de los cuales son del máximo interés como el de Lord (1980) y resultan recopilaciones muy adecuadas tanto de los desarrollos teóricos como de las aplicaciones, recogiendo hasta la fecha de modo muy adecuado el estado del campo. Además, el número de investigaciones crece de un modo enorme, con lo que reflejar aquí siquiera los artículos más relevantes excede del propósito del trabajo

En la tarea de hacer más accesible la teoría destaca R. K. Hambleton quien a partir de finales de los setenta empieza a publicar escritos introductorios y revisiones a la TRI. Así tenemos Hambleton (1979; 1983) y Hambleton y Cook (1977). La tarea de publicar manuales de tipo introductorio llevada a cabo por este autor culmina, por el momento, en 1985 firmando junto con Swaminathan el libro "Item Response Theory: Principles and Applications". Este libro presenta dos partes bien diferenciadas. La primera, dedicada a proveer los conceptos teóricos necesarios comprendería los capítulos del uno al siete. El capítulo uno, describe sumariamente la historia de la TRI. El segundo y el tercero presentan los supuestos y algunos de los principales modelos respectivamente. En el capítulo cuatro se trata el problema de la construcción de la escala de capacidad, mientras el capítulo seis trata del importante concepto de función de información y conceptos relacionados. Los capítulos cinco y siete giran

en torno a los principales métodos de estimación. La segunda parte del libro tendría un carácter más aplicado e incluiría los capítulos del ocho al catorce. Se revisan en estos siete capítulos los tópicos de: ajuste datos-modelo con algunos ejemplos (capítulos 8 y 9), igualación de tests (10), construcción y rediseño e tests (11), bancos de ítems (12), sesgos y tests a la medida (13), y un último capítulo (14) de consideraciones prácticas en el uso de modelos de TRI. El libro de fácil lectura repasa de modo elegante, riguroso e inteligible los desarrollos y aplicaciones más notables y actualmente en mayor uso, resultando muy recomendable como acercamiento a la materia.

Uno de los más importantes libros en TRI hasta el momento lo escribe Lord en 1980: "Applications of Item Response Theory to Practical Testing Problems". Este libro no sólo trata con las bases teóricas de la TRI sino que ofrece una visión amplia de las más importantes aplicaciones prácticas de esta teoría. A decir de muchos expertos es uno de los libros centrales de la TRI. El capítulo uno presenta una muy sumaria presentación de la teoría clásica de tests. El capítulo dos es una primera aproximación a la TRI presentando los supuestos y la función de respuesta al ítem, mientras el capítulo tres compara ambas ramas de la teoría de tests. El capítulo cuatro presenta el problema de la estimación. El capítulo cinco trata los conceptos de funciones de información del test y del ítem y los pesos óptimos de ponderación de los ítems, cerrando una primera parte dedicada a las bases de la teoría. La segunda parte incluye diversas aplicaciones entre las que Lord incluye el uso de la fórmula para la eficacia relativa (capítulo 6), el número de alternativas óptimo por ítem (7), tests de nivel flexible (8), tests en dos etapas y tests multinivel (9), tests a la medida (10), y tests de maestría (11). La tercera parte aparece bajo el epígrafe "problemas prácticos y otras aplicaciones", incluyendo un primer capítulo que insiste en el problema de la estimación de parámetros del ítem y la capacidad. El segundo capítulo de esta tercera parte (capítulo 13) trata sobre la igualación, mientras los capítulos catorce y quince plantean el problema de los sesgos y el problema de la puntuación cuando se tienen respuestas omitidas respectivamente. Por último, una cuarta parte del libro incluye dos capítulos sobre la estimación de la distribución de puntuaciones verdaderas en un test (capítulo 16) y en dos tests (17). El tono del libro es menos introductorio que el de Hambleton y Swaminathan (1985), pero presenta perfectamente la mayoría de los tópicos relevantes en la TRI actual siendo todavía de lectura imprescindible.

También de 1980 es el libro "Discrete Statistical Models with Social Science Applications" de E. B. Andersen. En este libro hay desarrollos de los modelos logísticos que están a la base de algunos procedimientos de calibración de ítems y estimación de parámetros en la TRI (Meliá, 1989).

Las actas del congreso celebrado con motivo del cincuenta aniversario del Australian Council for Educational Research fueron editadas en el año 1982 (D. Spearrit) con el título de "The improvement of Measurement in Education and Psychology". Pueden destacarse contribuciones como la de Keats que pone en relación la práctica de medición educativa desde la TRI y desde la teoría clásica. También es de interés el trabajo de Choppin sobre el uso de los modelos de rasgo latente en tests de habilidades cognitivas, o el de Andrich sobre el uso de estos mismos modelos para medición de actitudes.

En 1983 Charles L. Hulin, Fritz Drasgow y Charles K. Parsons publican el libro "Item Response Theory: Application to Psychological Measurement" (1983). Escrito en un nivel muy introductorio, intenta repasar todos los conceptos teóricos relevantes para después presentar algunas de las aplicaciones más importantes, como los tests a la medida o el estudio de sesgos. Especialmente relevante es el libro por la parte dedicada a la adecuación de la medición, en la que se presenta un resumen de la investigación realizada por los autores en ese campo. Se discuten en este capítulo cómo diversas violaciones en los supuestos de los modelos afectan al ajuste y precisión de la estimación. Otro libro de carácter muy introductorio es el de F. B. Baker "The Basics of Item Response Theory" publicado en 1985. Las nociones básicas sobre la teoría se presentan en los capítulos uno al seis, repasando entre otros los conceptos de curva característica del ítem, funciones de información, curva característica del test, métodos para la estimación de la capacidad de un solo sujeto. El resto del libro presenta sumariamente algunas aplicaciones. Se incluye en el libro un disco bien para Apple II o para PC que complementa el aprendizaje evitando cálculos iterativos muy tediosos y permitiendo ensayar ejemplos.

Howard Wainer y Samuel Messick, ambos del Educational Testing Service (ETS), publican en 1983 las actas del simposium celebrado por el ETS con motivo del setenta aniversario de F. M. Lord en 1982. El título es significativo "Principals of Modern Psychological Measurement". El libro no se circunscribe tan solo al campo de la TRI, si bien la parte más voluminosa del mismo está dedicada a este área. El libro presenta cinco partes desiguales en volumen entre las que se cuentan una sobre la teoría de respuesta al ítem y sus aplicaciones, otra sobre análisis factorial, y otra sobre otros modelos para medición psicológica. En la parte dedicada a la TRI sobresalen los trabajos de Green sobre tests a la medida, el de Bock sobre estimación bayesiana, la presentación de un nuevo modelo para datos en tablas de contingencia de Andersen, y el de Samejima sobre estimación máximo verosímil.

Todo esto en lo que se refiere a los libros que tratan más en profundidad los modelos logísticos de dos y tres parámetros y que prestan

menos atención a los distintos modelos de Rasch o lo consideran un submodelo del logístico de tres parámetros. Otros libros se han centrado justamente en los modelos de Rasch. Wright en 1980 reedita el libro de Rasch (1960). Este autor y sus colaboradores de Chicago han sido los que más han investigado en la aplicación de los modelos de un solo parámetro defendiendo su utilidad práctica y robustez frente a violaciones de sus supuestos (Muñiz, 1990).

Un importante libro para el acercamiento al campo de los modelos de Rasch es el de B. D. Wright y G. N. Master "Rating Scale Analysis" (1982), subtítulo "Rasch Measurement". De especial interés es el capítulo tres, donde se presentan los cinco submodelos del modelo de Rasch que más suerte han hecho en la literatura, además de una discusión sobre las propiedades de cada uno de ellos. El capítulo cuatro presenta una discusión en torno a cuatro procedimientos de estimación en el modelo de Rasch. Los capítulos seis, siete, ocho y nueve presentan tres investigaciones donde se ponen en práctica alguno de los modelos previamente explicados.

En 1984 Grijter y Van der Kamp escriben "Statistical Models in Psychological and Educational Testing". El libro trata de ofrecer un "tratamiento integrado del amplia área de la teoría de los tests mentales y de sus aplicaciones" desde las principales ramas de la psicometría. Las partes que se refieren directamente a TRI son la cuarta y la quinta. La parte IV presenta en dos capítulos los fundamentos teóricos de los modelos de TRI, el primer capítulo acerca de los modelos logísticos y de ojiva de dos y tres parámetros y el segundo íntegramente dedicado al modelo de Rasch. En la parte V el primer capítulo desglosa algunas aplicaciones como los tests a la medida o el estudio del sesgo de los ítems.

Del año 1988 es el libro de David Andrich "Rasch Models for Measurement". Este libro, publicado por SAGE, es un compendio de los modelos de Rasch más utilizados en breves páginas. Aun cuando el objetivo central son los modelos de Rasch no se circunscribe tan solo a ellos, contando con dos artículos introductorios muy generales, una serie de desarrollos más allá del modelo de un parámetro, y una interesante consideración sobre cuestiones de estimación y tests de ajuste (Meliá, 1989).

En cuanto a integraciones de los desarrollos teóricos en artículos de revisión, pueden resaltarse las revisiones del Annual Review of Psychology realizadas por Weiss y Davison (1981) y por Traub y Lam (1985). La primera revisión divide sus contenidos en teoría clásica de tests, alternativas a la teoría clásica y validez. La segunda revisión trata en profundidad la teoría de respuesta al ítem introduciéndola junto al análisis de clase latente como tipos de análisis de la estructura latente, el primero adecuado cuando la capacidad

se mide de modo continuo y el segundo cuando la capacidad se mide de modo discreto. Este trabajo es también relevante por su tratamiento introductorio de algunas de las aplicaciones de la TRI, en particular; la igualación de tests, los sesgos de los items, y los tests a la medida.

Aun cuando en los libros que hemos revisado anteriormente ya se trataban ampliamente las aplicaciones de la TRI, ahora veremos algunas aportaciones más centradas en la aplicación a la solución de problemas prácticos, tanto en libros como en algunos artículos seleccionados.

La revista *Applied Psychological Measurement* dedica un número especial en 1982 a la TRI que se titula "Advances in Item Response Theory and Applications", editado por R. K. Hambleton y W. J. Van der Linden. El primer artículo es de los editores y es introductorio a los temas que se tratan en los siete siguientes y se reúnen bajo tres tópicos los restantes trabajos: 1) Nuevos modelos (artículos de McDonald; Fischer y Formann; Mokken y Lewis). 2) Estimación de parámetros (Bock y Mislevy; Andersen). El artículo de Andersen compara los métodos de estimación de máxima verosimilitud condicional, directo y marginal. 3) Aplicaciones, con un artículo sobre el error típico de medida en la igualación (Lord) y otro sobre los tests a la medida que incluye una breve historia del tópico (Weiss)

Ya en 1983 es el libro que edita Weiss "New horizons in testing: Latent Trait Test Theory and Computerized Adaptive Testing". El libro contiene artículos como el de Levine y Drasgow sobre adecuación de la medida. Otro de Wainer sobre corrección del azar. Trabin y Weiss tratan el concepto de curva de respuesta de la persona. Thissen presenta un trabajo sobre el problema del tiempo para contestar a los items desde la TRI. Bock presenta una solución desde la TRI al problema del escalamiento mental. Samejima contribuye con un nuevo modelo, el modelo de información constante. Otras contribuciones se refieren a los tópicos de igualación, estimación de parámetros, y otros.

También de 1983 es el libro de R. K. Hambleton "Applications of Item Response Theory". Este libro se compone de 13 artículos escritos por diversos autores. Los cinco primeros artículos son una presentación general a la TRI, y los ocho restantes tratan propiamente aplicaciones. El trabajo de Swaminathan trata sobre estimación. El de Traub presenta diversos análisis sobre ajuste datos-modelo, concluyendo que el modelo de Rasch es en muchos casos inadecuado. Hambleton y Murray también tratan de la adecuación entre los datos y el modelo. Wigersky presenta el uso del programa LOGIST. La conclusión general a lo largo de los diversos trabajos es la superioridad del modelo de tres parámetros sobre el modelo de Rasch al ajustar los datos. En este libro de aplicaciones falta el tratamiento de los tests a la medida o tests adaptativos computerizados, considerándose

especialmente, el ajuste de los modelos a los datos. En el campo de la adecuación de los modelos a los datos y de la robustez de estos frente a violaciones podemos encontrar los trabajos de Drasgow (1982), Drasgow y Guertler (1987), Drasgow y Parsons (1983), Hambleton y Cook (1980), o Reckase (1979), además del libro de Hulin, Drasgow y Parsons (1983) ya comentado que resumía toda la investigación en este tópico de los autores, por lo demás especialistas en el campo

Sobre igualación, sólo un par de citas. Bejar y Wingersky publican un artículo titulado "A study of Pre-equating based on IRT" (1982) que plantea la posibilidad de pre-igualación mediante la TRI de un test, demostrando que los resultados son similares a los producidos mediante igualación a posteriori. Además, en 1982 se editan los trabajos presentados a la 'Test Equating Conference' organizada por el Educational Testing Service en 1980. Los editores son P. W. Holland y D. B. Rubin. El libro está compuesto de cuatro partes: análisis de los procedimientos de igualación usados por el ETS, métodos de igualación de la TRI, nuevos métodos para la igualación de tests, y en cuarto lugar tópicos relacionados y discusión. Buena parte de los procedimientos de igualación son tratados aquí en detalle. Las actas de este congreso tenían como objetivo declarado presentar el campo completo de la igualación y, sin embargo, métodos de igualación clásicos como el equipercantil no quedan reflejados en el texto. En cuanto a los tests a la medida o tests adaptativos computerizados, Isaac I. Bejar (1983) escribe "Achievement Testing. Recent Advances" que en parte se ocupa detalladamente de este tipo de aplicación. Y Van der Kamp, Langerak, y de Gruijter (1980) firman el libro "Psychometrics for Educational Debates" que recoge una selección de los trabajos presentados en el Third International Symposium on Educational Testing celebrado en Leyden. Los trabajos no tratan únicamente sobre la teoría de la respuesta al ítem, pero sí aparecen trabajos sobre estudio de sesgos y tests adaptativos computerizados realizados desde la TRI y firmados, entre otros, por Novick o Lord.

Los programas de ordenador más relevantes que aparecen en esta década son el BILOG (Mislevy y Bock, 1982) que puede utilizarse para la estimación de los modelos logísticos de uno y dos parámetros mediante estimación máximo verosímil marginal, PML (Gustafsson, 1980) que funciona para estimación con el modelo de un parámetro (tanto en ítems dicotómicos como multicategoriales) y mediante estimación máximo verosímil condicional, y MICROCAT (ASCAL Y RASCAL) (1982).

## 6. LA TRI EN ESPAÑA.

El desarrollo de la TRI en España es más tardío no produciéndose el mismo proceso de aplicación importante que si ha ocurrido, por ejemplo, en EEUU. No obstante, se han realizado un cierto número de contribuciones,

bien en forma de libros de introducción a la teoría, como de estudio de aplicaciones concretas en artículos o comunicaciones a congresos científicos.

Uno de los autores que más han contribuido a la difusión de la teoría de la respuesta al ítem en España es C. Santisteban. En un nivel de difusión introductoria tenemos los apuntes preparados en 1984 para cursos en la Complutense, que bajo el título "Los Modelos de Rasgo Latente en la Teoría de los Tests" recogen los principales modelos de teoría de la respuesta al ítem además de los modelos de la teoría fuerte de las puntuaciones verdaderas. También a este nivel tenemos cinco capítulos del libro "Psicometría. Teoría y práctica en la construcción de tests" de 1990. La zona de este libro dedicada a la Teoría de Respuesta al Ítem abarca desde el capítulo 9 hasta el 14. El capítulo 9 incluye aspectos introductorios y generales de la TRI: breve historia, supuestos, puntuaciones verdaderas y escala de capacidad,... El capítulo 11 presenta los modelos logísticos, los modelos de ojiva normal, y otros modelos para respuestas no dicotómicas como los modelos de respuesta continua y los modelos de respuesta nominal. El capítulo 12 se dedica a la estimación. Se presentan los métodos de máxima verosimilitud condicional e incondicional y los métodos de Bayes o Bayesianos. El capítulo 13 presenta las funciones de información del ítem y del test, concepto central dentro de la teoría para propósitos prácticos de construcción de tests. Otros conceptos relacionados que se presentan en el mismo capítulo son las ponderaciones óptimas de los ítems y las funciones de eficiencia relativa. El capítulo 14 es una bibliografía comentada y ordenada cronológicamente de interés para el que desee profundizar en el conocimiento de la TRI. El capítulo 15 trata sobre el análisis de ítems y la construcción de tests tanto desde la teoría clásica como desde la teoría de respuesta al ítem.

Esta misma autora presenta un número de investigaciones concretas en aspectos teóricos y de aplicación de la teoría, como pueden ser los dedicados a modelos y estimación (Santisteban 1985, 1986, 1988). A destacar la comunicación de 1986 que presenta un repaso a los métodos de estimación de parámetros en TRI, y el artículo publicado en *Psicológica* en 1988 que provee de una aproximación Bayesiana al problema de la estimación de la capacidad desde un Modelo de Rasch.

Otro autor que ha trabajado tanto en la divulgación como en la aplicación de la TRI en nuestro país es J. Muñiz. Exponente de la vertiente introductoria es su libro "Teoría de Respuesta a los Ítems". El libro consta de nueve capítulos, los dos primeros de carácter introductorio. El tercer capítulo se dedica a la presentación de los supuestos en que se basa el funcionamiento de la TRI, de unidimensionalidad, independencia local y curvas características del ítem; presentándose comprensivamente la noción



de los parámetros correspondientes a los modelos más conocidos de TRI. El capítulo cuatro presenta los modelos logísticos y de ojiva normal de un, dos, tres y cuatro parámetros. El autor se cñe a estos sin olvidar reseñar que existen otros posibles. El capítulo cinco, comprobación de los modelos, presenta modos prácticos de elegir el tipo de modelo más adecuado, la estimación de los parámetros del ítem y de la capacidad, y la comprobación del ajuste entre el modelo y los datos. El método de estimación presentado es el de la máxima verosimilitud. Se remite al lector a otros textos para la explicación de otros métodos de estimación (v. gr. Bayes). Cierra el capítulo una interesante reproducción de un estudio de simulación publicado por el autor y otros en la revista *Anuario de Psicología* (1989), encaminado a la comprobación de la robustez del modelo de un parámetro frente a violaciones de sus supuestos. Las consideraciones al respecto de la escala de medida de la capacidad vienen tratadas en el capítulo seis. Se presentan las transformaciones lineales de la capacidad (transformaciones admisibles), un caso de transformaciones no lineales (escala en logits), y otras. El capítulo ocho presenta básicamente los conceptos de curva característica del test y curva característica de la persona. La función de información y algunas cuestiones relacionadas son el tema del capítulo siete. En concreto se tratan los tópicos de función de información, información máxima, ponderaciones óptimas de los ítems, y eficiencia relativa. De carácter práctico es el capítulo nueve que presenta las siguientes aplicaciones de la TRI: bancos de ítems, equiparación, sesgo, tests a medida, y tests referidos al criterio. Cierra el libro un breve capítulo dedicado a recapitulación y consideraciones finales.

En 1990 aparece el libro "La construcción de la Psicometría como ciencia teórica y aplicada" de J.L. Meliá, donde se presenta una teoría integrada sobre el campo psicométrico y una revisión general de toda la Psicometría. El libro considera a la TRI como una de las grandes ramas de la Psicometría y dedica el capítulo 7 a efectuar una introducción histórica y conceptual a sus tópicos principales, introduciendo síntesis ordenadoras de los principales modelos y una exposición didáctica de los conceptos principales. En la cronología de esta obra pueden encontrarse comentadas los principales avances y aportaciones al campo, principalmente libros y artículos de revisión, hasta la fecha de su publicación.

Otro libro de carácter introductorio a la TRI orientado básicamente a la docencia aparece en 1991 firmado por Tomás, Oliver y Meliá. El libro consta de siete capítulos. El primero, fundamentos de la TRI, presenta los supuestos de la teoría y el concepto de función de respuesta al ítem. El segundo capítulo recoge los principales modelos de la TRI tanto dicotómicos como politómicos o continuos, con especial interés en los logísticos binarios. El tercer capítulo trata sobre la escala del rasgo latente y las transformaciones admisibles de ésta, la curva característica del test y las puntuaciones verdaderas. La estimación de parámetros desde el método de

máxima verosimilitud condicional y máxima verosimilitud conjunta queda recogida en el capítulo cuatro. La función de información, el error típico de medida, las ponderaciones óptimas, y la eficiencia relativa se tratan en el cinco. El capítulo siete compara la teoría clásica de tests y la TRI tanto desde los conceptos teóricos como desde las aplicaciones de ambas teorías. Cierra el libro un capítulo de M. D. Sancerni que recoge las más importantes aplicaciones de la TRI. Se repasan en relativa profundidad los tópicos de rediseño de tests, tests de nivel flexible, tests en dos etapas, tests en múltiples etapas, tests ramificados, tests de maestría, igualación, y estudio de sesgos.

Al tiempo que ultimábamos este artículo se publicó en el Anuario de Psicología uno de similar contenido (Muñiz y Hambleton, 1992). El trabajo, a la vez introductorio y de revisión histórica resulta muy adecuado para los psicómetras interesados en profundizar a través de las numerosas y relevantes citas que recogen las principales contribuciones, tanto teóricas como de aplicación, de la TRI a la psicometría actual. Al mismo tiempo puede servir de introducción mínima a los principales conceptos que maneja la teoría, además de recuerdo del marco psicométrico clásico en el que apareció allá por los 40-50 esta vigorosa teoría psicométrica.

En cuanto a desarrollos teóricos destacar los trabajos de De la Fuente y Pérez-Meléndez (1987) y Padilla y Pérez-Meléndez (1991). Los primeros presentan en la revista *Psicológica* un desarrollo teórico conectando el concepto de función de información tal y como ha sido usado en TRI con el concepto de información definido tradicionalmente en estadística matemática. Por su parte Padilla y Pérez-Meléndez a partir del modelo de "rating" de Rasch presentan en el III Congreso Nacional de Psicología Social una comunicación acerca de sus implicaciones como modelo aprovechable para la medición de actitudes y su conexión con las tradiciones de Thurstone y Likert.

Por último, y representativas de estudios de aplicación más específicos tenemos tres comunicaciones presentadas a las sucesivas Conferencias Españolas de Biometría (1989) que junto con el grupo español de la Biometric Society, han jugado un importante papel como lugar de encuentro y debate de psicómetras y matemáticos interesados en el tema.

Entre las comunicaciones podemos destacar la presentada por S. Herrando donde se discute una solución a los problemas de estimación que se producen en los tests a la medida al utilizar los estimadores máximo verosímiles cuando aparecen puntuaciones cero o perfectas. Artés, Herrando y Renom presentan una comparación entre estimaciones obtenidas mediante los programas informáticos BICAL, RASCAL y ASCAL

bajo los modelos logísticos de uno, dos y tres parámetros. El último trabajo, firmado por Gómez, Artés y Deurnal estudia el efecto de cambios en el número de alternativas de respuesta sobre las estimaciones de un modelo de Rasch.

Los apuntes anteriores sobre la TRI en España son, obviamente, incompletos, debido a las dificultades de comunicación científica que en nuestro país, como fruto de la ausencia de canales e índices adecuados, todavía subsisten.

## 7. CONCLUSIONES.

La teoría de la respuesta al ítem ha tenido un desarrollo espectacular desde que en los años cincuenta empezara a gestarse. Como hemos visto en este trabajo, la evolución histórica ha pasado por diversas etapas. En un último intento aún más sintetizador podríamos decir que dos momentos se perfilan claramente en estas cuatro décadas revisadas. En un primer momento, la gran mayoría de trabajos tienen un fuerte carácter teórico, de desarrollo matemático de la teoría. Podemos decir que este periodo dura, al menos, hasta 1968 en que aparece el libro de Lord y Novick "Statistical Theories of Mental Test Scores". Un segundo momento a partir de aquí incluiría cada vez mayor número de aplicaciones. Este último proceso se ve propiciado fundamentalmente por dos motivos. Por un lado, hay un cuerpo teórico-matemático importante que sostiene la teoría. Por otro, se empieza a disponer de programas informáticos para implementar la teoría, que resulta del todo inmanejable sin ellos. Por supuesto, en la realidad, los límites no son ni mucho menos claros. Importantes desarrollos teóricos se han dado más allá del 68 y antes de esta fecha ya se implementaba la TRI y se estudiaba su aplicación. Estos tipos de clasificaciones son siempre hasta cierto punto arbitrarias y se han utilizado aquí tan solo con un énfasis clarificador. Sin duda la TRI es un cuerpo de Teoría y de Aplicación Psicométrica que goza de buena salud, acaparando una parte principal de los esfuerzos de investigación en Psicometría. En España, todavía hay un énfasis importante en la divulgación de la Teoría, que paulatinamente habrá de abrir paso a la aplicación y avances en la misma.

## BIBLIOGRAFIA

- ANDERSEN, E. B. (1973). Goodness of Fit Test for the Rasch Model. *Psychometrika*, Vol. 38, 123-140
- ANDERSEN, E. B. (1977) Sufficient statistics and latent trait models. *Psychometrika*, Vol. 42, 69-81

- ANDERSEN, E. B. (1980). *Discrete Statistical Models with Social Science Applications*. North Holland Publishing Company, Amsterdam.
- ANDRICH, D. (1978). Relationships between the Thurstone and Rasch approaches to item scaling. *Applied Psychological Measurement*, Vol 2, pp: 451-462.
- ANDRICH, D. (1988). *Rasch models for measurement*, Beverly Hills, California, SAGE.
- ARTES, M.; HERRANDO, S. y RENOM, J. (1989). Comparación entre estimaciones de parámetros obtenidas a partir de los programas BICAL, RASCAL, ASCAL. *Actas de la II Conferencia Española de Biometría*. Segovia.
- BAKER, F. B. (1983). Comparison of Ability Metrics Obtained under Two Latent Trait Procedures. *Applied Psychological Measurement*, Vol 7, nº 1, 97-110.
- BAKER, F. B. (1985). *The Basics of Item Response Theory*. Heinemann Educational Books, Portsmouth.
- BARTON, M.A. y LORD, F.M. (1981). An upper asymptote for the three-parameter logistic item-response model. *Research Bulletin* 81-20. Princeton, NJ: Educational Testing Service.
- BEJAR, I. I. (1983). *Achievement Testing. Recent Advances*. SAGE, Beverly Hills.
- BEJAR, I. I.; WEISS, D. J. (1979). Computers programs for scoring test data with item characteristic curve models. *Journal of Applied Psychology*, Vol. 62, 155-57.
- BEJAR, I. I. y WINGERSKY, M. S. (1982). A Study of Preequating Based on ITeM Response Theory. *Applied Psychological Measurement*, Vol. 6, nº 3, 309-325.
- BERK, R. (Ed.): *Handbook of Methods for detecting test bias*. The Jones Hopkins University Press. Baltimore, M.D.
- BINET, A. y SIMON, T. H. (1916). *The Development of Intelligence in Young Children*. Vineland, NJ: The TRaining School.
- BIRNBAUM, A. (1957). Efficient design and use of tests of a mental ability for various decisions making problems. Series report No. 58-16. Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas.
- BIRNBAUM, A. (1958). On the estimation of mental ability. Series report No. 15. Project No. 7755-23, USAF School of Aviation Medicine, Randolph Air Force Base, Texas.
- BIRNBAUM, A. (1961). A unified Theory of estimation, I. *Annals of Mathematical Statistics*. Vol. 32, 112-135.
- BIRNBAUM, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. En LORD, F. M.; NOVICK, M. N. (1968). *"Statistical Theories of Mental Test Scores"*. Addison-Wesley; Reading, Massachusetts

- BIRNBAUM, A. (1969). Statistical Theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology*, 6, 258-276.
- BOCK, R. D.; WOOD, R. (1971). Test Theory. *Annual Review of Psychology*, Vol 22, 193-224.
- BRODGEN, H. (1946). Variations in Test Validity with Variation in the Distribution of Item Difficulties, Number of Items, and Degree of the Intercorrelations. *Psychometrika*, Vol. 11, nº 4, 197-214.
- BRODGEN, H. (1977). The Rasch Model, the law of comparative judgment and additive conjoint measurement. *Psychometrika*, Vol 42, pp: 631-34.
- COHEN, L. (1979). Approximate methods for parameters estimates in the Rasch model. *Br. J. Math. Stat. Psychol.*, Vol 32, pp: 113-20.
- DINERO, T. E.; HAERTL, E. (1977). Applicability of the Rasch Model with varying item discriminations. *Applied Psychological Measurement*, Vol. 1, pp: 581-92.
- DRASGOW, F. (1982). Choice of test for apropiateness measurement. *Applied Psychological Measurement*, Vol. 6.
- DRASGOW, F. y GUERTLER, E. (1987). A Decision Theoretic Approach to the Use of Apropiateness Measurement for Detecting Invalid Test and Scale Scores. *Journal of Applied Psychology*, Vol. 72, nº1, 10-18.
- DRASGOW, F. y PARSONS, C. K. (1983). Applications of Unidimensional Item Response Theory Models to Multidimensional Data. *Applied Psychological Measurement*, Vol. 7, nº 2.
- GOMEZ, J. ; ARTES, M. y DEUMAL, E. (1989). Efecto del número de rangos de la escala de medida sobre la calibración de ítems y sujetos mediante credit. *Actas de la II Conferencia Española de Biometría*. Segovia.
- GREEN, B. F. (1954). Attitude Measurement. En G. Lindzey (Ed.), *Handbook of Social Psychology*, Reading Mass.: Adisson-Wesley Publishing.
- GRUIJTER, D. N. M. y Van der KAMP, L. J. Th. (1984). *Statistical Models in Psychological and Educational Testing*. Swets Zeitlinger. Lisse.
- GUSTAFSON, J. E. (1979). The Rasch Model in vertical equating of tests: A critique of Slinde and Linn. *Journal of Educational Measurement*, Vol. 16, pp: 153-58.
- GUTTMAN, L. (1950). Capítulos 2, 3, 6, 8. En STOUFFER et al (Eds) *Measurement and Prediction*. Princeton, N. J.: Princeton University Press.
- HAMBLETON, R. K. (1979). Latent Trait Models and their Applications. *Methodological Developments*, Vol. 4. San Francisco: Jossey-Bass. En Traub, R. (Ed). *New Directions for Testing and Measurement*.

- HAMBLETON, R. K. (1983). Applications of Item Response Models to Criterion-Referenced Assessment. *Applied Psychological Measurement*, Vol 7, nº 1, 33-44.
- HAMBLETON, R. K. (1983). *Applications of Item Response Theory*. Educational Research Institute of British Columbia, Vancouver.
- HAMBLETON, R. K.; COOK, L. L. (1977). Latent-Trait Models and Their Use in the Analysis of Educational Test Data. *Journal of Educational Measurement*, Vol 14, 75-95.
- HAMBLETON, R. K.; COOK, L. L. (1980). The robustness of latent trait models and effects of test length and sample size on the precision of ability estimates. En Weiss (1980), pp: 349-64.
- HAMBLETON, R.K. Y ROVINELLI, R.A. (1973). A fortran IV program for generating examinee response data from logistic test models. *Behavioral Science*, 17, 73-74.
- HAMBLETON, R.K.y SWAMINATHAN, H. (1985). "Item Response Theory: Principles and Applications". Kluwer-Nijhoff Publishing, Boston.
- HAMBLETON, R. K. y Van der LINDEN, W. J. (1982). Advances in Item Response Theory and Applications. *Applied Psychological Measurement*, Vol 6, nº 4
- HERRANDO, S. (1989). Tests adaptativos computerizados: una sencilla solución al problema de la estimación con puntuaciones perfecta y cero. *Actas de la II Conferencia Española de Biometría*. Segovia.
- HOLAND, P. W. y RUBIN, D. B (1982) *Test Equating*. Academic Press, New York.
- HULIN, C. L.; DRASGOW, F. y PARSONS, C.K. (1983). *Item Response Theory. Application to psychological measurement*. Homewood, Illinois, Dow Jones-Irvin.
- JENSEMA, C. J. (1976). A simple technique for estimating latent trait mental test parameters. *Educational and Psychological Measurement*, Vol. 36, pp: 705-15.
- KEARNS, J.; MEREDITH, W. (1975). Methods for evaluating Bayes point estimates of latent trait scores. *Psychometrika*, Vol 40, 373-94.
- KEATS, J. A. (1974). Applications of projective transformations to test theory. *Psychometrika*, Vol 39, pp: 359-60.
- LAZARSELD. P.F. (1950). Capítulos 10 y 11. En STOUFFER et al (Eds) *Measurement and Prediction*. Princeton, N. J.: Princeton University Press.
- LAZARSELD. P.F. (1959). Latent Structure Analysis. En S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3). New York: McGraw-Hill, 1959, pp:476-542.
- LAZARSELD. P.F. y HENRY, N. W (1968). *Latent Structure Analysis*. New York: Houghton Mifflin.
- LAWLEY, D. N. (1943). On Problems Connected with Item Selection and Test Construction. *Proceedings of the Royal Society of Edimburg*, 61, 273-287.

- LAWLEY, D. N. (1944). The Factorial Analysis of Multiple Item Tests. *Proceedings of the Royal Society of Edimburg*, 62-A, 74-82.
- LEWIS, Ch. (1986). Test Theory and Psychometrika: The Past Twenty-five Years. *Psychometrika*. Vol. 51, nº1, 11-22.
- LORD, F. M., (1952). A theory of tests scores. *Psychometric Monographs*, 1952, Nº 7.
- LORD, F. M., (1953). An application of confidence intervals and of maximum likelihood to the estimation of an examinee ability. *Psychometrika*, Vol. 18, 57-76.
- LORD, F. M.; NOVICK, M. N. (1968). "Statistical Theories of Mental Test Scores". Addison-Wesley; Reading, Massachusetts.
- LORD, F. M. (1980). "Applications of Item Response Theory to practical testing problems". Lawrence Earlbaum Associates, Hillsdale, New Jersey.
- LORD, F. M.; STOCKING, M. L. (1988). Item Response Theory. En KEEVES, J. P. (1988). "Educational Research, methodology, and measurement. An international handbook". Pergamon Press, Oxford.
- LUMSDEN, J. (1976). Test Theory. En M. R. Rosenzweig y L. W. Porter (Eds), *Annual Review of Psychology*. Palo Alto, California: Annual Reviews Inc.
- LUMSDEN, J. (1980). Variations on a theme by Thurstone. *Applied Psychological measurement*, Vol. 4, pp: 1-7
- McDONALD, R. P. (1967). Non-linear factor analysis. *Psychometric Monographs*, Nº 15.
- MELIA, J. L. (1989). *Psicometria* Proyecto docente. Universitat de València.
- MELIA, J. L. (1990). *La construcción de la Psicometría como ciencia teórica y aplicada*. Ed. C. Serrano, València.
- MISLEVY, R.J. y BOCK, R.D. (1984). *BILOG versión 2.2. Item analysis and test scoring with binary logistic models*. Mooresville, I.N. Scientific Software.
- MUÑIZ, J. (1990). *Teoría de Respuesta a los Items, un nuevo enfoque en la evolución psicológica y educativa*. Pirámide, Madrid.
- MUÑIZ, J. Y HAMBLETON, R.K. (1992). Cincuenta años de Teoría de Respuesta a los Items. *Anuario de Psicología*, nº 52.
- PADILLA, J.L. y PEREZ, C. (1990). El modelo de Rasch y las aproximaciones tradicionales a la medición de actitudes. *Libro de Comunicaciones del III Congreso Nacional de Psicología Social*, (I), 416-427.
- RASCH, G. (1960). *Probabilistics Models for Some Intelligence and Attainment Tests*. Denmarks paedagogiske Institut. Copenhagen.
- RECKASE, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *J. Educ. Stat.*, Vol 4, pp: 207-30.

- RENTZ, F. R.; BASHAW, W. L. (1977). The National Reference Scale for Reading: An application of the Rasch Model. *Journal of Educational Measurement*, Vol. 14, 161-79.
- RICHARDSON, M. W. (1936). The relationship between Difficulty and the Differential Validity of the Test. *Psychometrika*, 1, 33-49.
- SAMEJIMA, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, N° 17.
- SAMEJIMA, F. (1973). Homogeneous case of the continuous response model. *Psychometrika*, Vol. 38, pp:203-219.
- SANTISTEBAN, C. (1977). Sobre Procesos Estocásticos de Ferguson, Neutros y de Cola Libre. *OJO!!!!*
- SANTISTEBAN, C. (1982). *Concepto, Método. Fuentes y Programa de Psicología Matemática II*. Madrid.
- SANTISTEBAN, C. (1984). *Los Modelos de Rasgo Latente en la Teoría de los Tests*. Universidad Complutense de Madrid.
- SANTISTEBAN, C. (1985). Modelos probabilísticos para la evaluación de las diferencias individuales. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales*, 79, 361-365.
- SANTISTEBAN, C. (1986). Estimación de parámetros en los modelos de Rasgo Latente. *Actas de la I Conferencia Española de Biometría*. Granada.
- SANTISTEBAN, C. (1988). The Dirichlet process and its application to latent trait models. *Psicológica*, 9. (2). 149-158.
- SANTISTEBAN, C. (1990). *Psicometría: teoría y práctica en la construcción de ítems*. Madrid, Norma.
- SANTISTEBAN, C. (1990b). *Curso sobre Teoría de Tests*. Valencia.
- SLINDE, J. A.; LINN, R. L. (1977). An exploration of the adequacy of the Rasch Model for the problems vertical equating. *Journal of Educational Measurement*, Vol. 15, 23-25.
- SLINDE, J. A.; LINN, R. L. (1979). A note on vertical equating via the Rasch Model for groups of quite different ability and tests of quite different difficulty. *Journal of Educational Measurement*, Vol. 16, 159-65.
- SPEARIT, D. (Ed) (1982). *The Improvement of Measurement in Education and Psychology*. Proceedings of the Invitational Seminar for the Fiftieth Anniversary of the Australian Council of Educational Research, Mearlbourne.
- TOMAS, J.M., OLIVER, A. y MELIA, J.L. (1992) Teoría de la Respuesta al ítem. Fundamentos. Modelos y Aplicaciones. Ed. Cristobal Serrano, Valencia
- TORGERSON, W.S. (1958) *Theory and methods of scaling*. New York: John Wiley & Sons
- TRAUB, R. E.; LAM, R. (1985). Latent Structure and Item Sampling Models for Testing. *Annual Review of Psychology*, Vol. 36, pp: 19-48.
- TUCKER, L.R. (1946). Maximum Validity of a Test with Equivalent Items. *Psychometrika*, 11, 1-13.



- TUCKER, L. R. (1951). Academic ability test. *Research Memorandum* 51-17. Princeton, N. J.: Educational Testing Service.
- URRY, V.W. (1977). Tailored testing: A success application of latent trait theory. *Journal of Educational Measurement*, 14, 181-196.
- VAN der KAMP, J. T.; LANGERAK, W. F.; De GRUIJTER, P. N. M. eds (1980). *Psychometrics for Educational Debates*. Wiley, New York.
- VAN der LINDEN, W. J. (1986). The Changing Conception of Measurement in Education and Psychology. *Applied Psychological Measurement*, Vol. 10, nº 4, 325-332.
- WAINER, H. y MESSICK, S. (1983). *Principals of Modern Psychological Measurement*. Lawrence Earlbaum Associated Publishers, Hillsdale, New Jersey.
- WEISS, D. J. (1983). *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing*. Academic Press, New York.
- WEISS, D. J. y DAVIDSON, M. I. (1981). Test Theory and Methods. *Annual Review of Psychology*, Vol. 32, 629-658.
- WOOD, R.L. y LORD, F.M. (1976). A user's guide to LOGIST. Research Memorandum, 76-4. Princeton, New Jersey. Educational Testing Service.
- WRIGHT, B. D. y MASTERS, G. N. (1982) *Rating Scale Analysis*. Mesa Press, Chicago, Illinois.
- WRIGHT, B.D. y MEAD, R.J. (1976). BICAL calibrating rating scales with the Rasch model. Research Memorandum N° 23. Statistical Laboratory. Department of Education. Chicago. University of Chicago.
- WRIGHT, B. D.; PANCHAPAKESAN, N. (1969). A procedure for samplefree item analysis. *Educational and Psychological Measurement*, 29, pp: 23-48.