

7. APROXIMACION A LA TEORIA DE LA RESPUESTA AL ITEM.

1. Introducción.

La psicometría, como campo referido a la medición psicológica en toda su amplitud, se ha desarrollado en dos grandes ramas: la teoría de la medición y el escalamiento psicológico, por una parte, y la teoría de los tests, por otra. La teoría de los tests, a su vez, como hemos visto, puede subdividirse en dos grandes cuerpos teóricos: la teoría clásica de los tests (TCT) y la teoría de la respuesta al ítem (TRI).

La teoría psicométrica clásica a pesar de ser históricamente anterior, no puede considerarse sin más caduca o "superada" por la Teoría de la Respuesta al Ítem. La TRI apenas hace uso de los resultados de aquella, pero tampoco la contradice o invalida.

Puede afirmarse que la TRI se inició propiamente en 1.950 con Lazarsfeld, y tiene sus principales desarrollos en los trabajos de Rasch, Lord y Hambleton, para citar algunos autores muy destacados.

Supuestos de la TRI. La TRI hace el supuesto de que la respuesta de un sujeto a un ítem depende de dos cosas.

Primero, de la posición del sujeto en un rasgo latente único -para modelos unidimensionales-, generalmente simbolizado por θ , considerado como un continuo en el que se pretende medir. La posición de un sujeto en θ significa su nivel de 'capacidad' o su nivel de 'aptitud', por ejemplo su nivel de 'inteligencia' en un test de inteligencia o su nivel de conocimientos en un test de rendimiento. En términos más generales la posición de un sujeto en θ significa su nivel en el rasgo medido, independientemente de la naturaleza psicológica del mismo.

Segundo, de las características psicométricas del ítem que describen la relación del ítem con el rasgo latente subyacente. Esas características incluyen, por ejemplo, la dificultad del ítem.

De ese modo la probabilidad de obtener determinada respuesta (usualmente, para ítems con respuesta

correcta, la probabilidad de obtener la respuesta correcta) a un ítem depende a la vez del nivel del sujeto en la característica que se está midiendo, por una parte, y de las propiedades del ítem por otra.

Función de respuesta al ítem. La teoría de la respuesta al ítem se funda sobre el supuesto de que se puede describir mediante una función matemática la relación existente entre la posición de un sujeto en θ y su probabilidad de dar una respuesta determinada a un ítem. La función matemática que relaciona el nivel en θ con la probabilidad de dar una respuesta determinada a un ítem se denomina *curva característica del ítem*, curva de huella del ítem o función de respuesta al ítem. Se han usado diversas funciones matemáticas para describir esa relación. Entre ellas pueden destacarse la función de ojiva normal descrita en 1.943 por Lawley y la función logística, que es la de más uso.

La teoría de la respuesta al ítem tiene en las obras de Lazarsfeld (1.950), Rasch (1.960), Birnbaum (1.968), Lord (1.980), Wainer y Messick (1.983) y De Gruijter y Van der Kamp (1.984) algunos de sus principales trabajos. Santisteban ha efectuado una exposición de los modelos de rasgo latente (1984) y de la TRI (1990).

2. Una perspectiva histórica de la teoría de la respuesta al ítem.

Se ha afirmado que la teoría de la respuesta al ítem surgió en los años cincuenta y sesenta como una reacción a la teoría clásica (Van der Linden, 1.986). Hambleton y Van der Linden (1.982) han justificado la teoría de la respuesta al ítem en las deficiencias y dificultades que la teoría clásica de test no podía resolver satisfactoriamente. Esta idea podría formularse más bien positivamente considerando que la teoría de la respuesta al ítem ha alcanzado un alto grado de desarrollo por sus propias virtudes para tratar con problemas principales de teoría psicométrica. A diferencia de la teoría clásica, que tiene su foco de atención en los tests, la TRI se centra, en primer lugar, como hemos visto, sobre los ítems individuales y las respuestas de los sujetos a los mismos.

En realidad, los orígenes de la teoría de la respuesta al ítem deben retrotraerse hasta 1.943, fecha en la que Lawley realizó una primera descripción del modelo de ojiva normal como una función indicada para expresar la relación entre una

dimensión latente de capacidad, tomada como continua, con la probabilidad de respuesta correcta a un ítem. Sin embargo, el primer trabajo de entidad que recibió una cierta atención fue el capítulo de Lazarsfeld de 1.950 "Latent structure theory" que aportó la denominación general de esta teoría durante los treinta años siguientes y el concepto de "*ítem trace line*" o curva característica de ítem que ocupa un lugar central en la teoría. El capítulo de Lazarsfeld a veces ha sido considerado como inaugural de la teoría de la respuesta al ítem.

No deja de ser llamativo que en este año de la mitad del siglo XX se produzcan a la vez la madura obra de Gulliksen, que resume el medio siglo anterior de trabajos en teoría clásica, y la seminal de Lazarsfeld, que abría el que mucho después llegaría a ser principal cuerpo teórico alternativo al de la teoría clásica. Al año siguiente del capítulo de Lazarsfeld fue leída la tesis "A theory of test scores" de F.M. Lord. La tesis de Lord, publicada como la monografía número siete de Psychometrika al año siguiente, constituye también uno de los primeros trabajos en teoría de la respuesta al ítem.

En 1.960 se produciría la publicación de la obra de Rasch "*Probabilistic models for some intelligence and attainment tests*" que abría el camino hacia los modelos un solo parámetro y sería el punto de arranque de una extensa investigación todavía muy activa. En el año 1.968 Lazarsfeld y Henri extendían la consideración de los modelos de clase latente y sobre todo Lord y Novick en su principal obra "*Statistical theories of mental tests scores*" ofrecen una presentación bien estructurada del modelo de ojiva normal y de los modelos logísticos que han llegado a ser los más estudiados y utilizados.

Desde un punto de vista histórico, la teoría de la respuesta al ítem no puede ser tratada independientemente de la consideración de la obra de Frederic Mather Lord, sin duda una de las figuras más relevantes de la Psicometría. En torno a su obra se pueden articular los desarrollos clásicos, que sintetiza y desarrolla con elegancia y claridad, y todos los avances teóricos que implican los modelos de la teoría del rasgo latente, mejor denominados de la teoría de la respuesta al ítem. F.M. Lord nació el 12 de Noviembre de 1.912 en Hannover, New Hampshire. Se graduó en Sociología en 1.936 y obtuvo el grado de master en Psicología Educativa en la Universidad de Minnesota en 1.943. Desde 1.941 trabajó en la Civil Service Commission, bajo la dirección de Marion Richardson y con colegas como B.J. Winer, en una tarea que le suponía

principalmente escribir ítems de tests. En 1.944 publica sus tres primeros trabajos: uno acerca del efecto de variar el número de alternativas de los ítems sobre la fiabilidad, el segundo relativo al cálculo de las correlaciones, y el tercero acerca de una aplicación específica de un test. En ese año también ingresa en el Educational Testing Service, en el que después de pasar por diversos cargos y tareas menores, accede, en 1.949, a la División de Investigación, donde desarrollará el resto de su carrera. Además, fue profesor visitante en diversos periodos de las universidades de Princeton, Wisconsin y Pennsylvania.

Lord se doctoró en psicología en la Universidad de Princeton en 1.951, con una tesis titulada "A Theory of Test Scores". La tesis, publicada como una monografía de "Psychometrika" en 1.952, contenía los resultados de un trabajo que había comenzado en 1.941 acerca de un nuevo conjunto de modelos que incluía algunos aportados por Thurstone, Lawley y Guttman, y que se revelaría capaz de incluir como un caso particular los que daría a la luz Rasch en 1.960. Ese nuevo conjunto de modelos es lo que después se ha conocido como Teoría de la Respuesta al Ítem (TRI). La tesis de Lord junto con el trabajo de Gulliksen "*Theory of Mental Tests*" constituyeron a juicio de Wainer "las bases de la mayor parte de la metodología del Educational Testing Service durante los siguientes treinta años" (1.983). Sin embargo, la principal contribución de Lord todavía no se había producido.

El hecho más destacable del año 1.960 desde el punto de vista de la teoría de los tests es la publicación de "*Probabilistic Models for Some Intelligence and Attainment Tests*" por G. Rasch. En la monografía, que alcanzará una importancia esencial para la evolución de la psicometría, se desarrollan dos modelos, el primero a propósito de los errores cometidos en lectura oral, el segundo relativo a la velocidad en un test de lectura. Desde luego los temas concretos que permitieron el desarrollo de lo que después se ha denominado el "modelo de Rasch" quedaron rápidamente trascendidos. Rasch presupone una (sola) dimensión latente a los ítems y postula una función de relación de un solo parámetro. Las aportaciones de la monografía de Rasch podían considerarse un desarrollo particular de la teoría de la estructura latente de Lazarsfeld, o, de otro modo, un caso particular -con características de una importancia esencial- de teoría de la respuesta al ítem.

En 1.968 Lord y M.R. Novick producen "*Statistical Theories of Mental Test Scores*", la que ha fue calificada como obra central de la Psicometría, o todavía más, como aseveran Bock y Wood en una revisión de 1.971, "el suceso señal de la década". El libro reflejaba la investigación acerca de test mentales producida desde la publicación de la tesis de Lord en 1.952, significando una síntesis y culminación de la teoría anterior, y abriendo a la vez caminos hacia nuevos desarrollos teóricos, entre ellos los aportados por los cuatro capítulos de Allan Birnbaum acerca de "Modelos de Rasgo Latente". Desde el punto de vista de la teoría clásica de los tests, la obra constituye un acontecimiento que puede considerarse sucesor de la obra de Gulliksen "*Theory of Mental Test*" aparecida en 1.950. Desde el punto de vista de la teoría de la respuesta al ítem constituye también una pieza fundamental. La obra sistematiza y somete a estudio riguroso una serie de modelos que ya habían sido desarrollados, cumpliendo un papel de recopilación y ordenación del campo, pero a la vez contiene innovaciones y aportaciones novedosas en TRI, buena parte de ellas debidas a Birnbaum.

La tercera gran contribución de Lord vendría doce años después, en 1.980, con la publicación de "*Applications on ítem response theory to practical testing problems*", una visión global no solo de las aplicaciones, sino también del cuerpo teórico de la teoría de la respuesta al ítem. El capítulo uno de esta obra es un ejemplo magistral acerca de como puede ser presentada la teoría clásica de los tests en solo ¡cuatro páginas! y un apéndice de otras tres, con una absoluta elegancia matemática que distingue claramente los verdaderos supuestos simples de la teoría de las puntuaciones verdaderas y la formulación de los tests paralelos. Los capítulos dos a cinco realizan una presentación estructurada de la teoría de la respuesta al ítem. Los capítulos seis a quince presentan un conjunto de aplicaciones y resolución de problemas prácticos de la teoría de la respuesta al ítem, y los últimos dos capítulos se dedican a la estimación de las distribuciones de puntuaciones verdaderas.

Con motivo del setenta cumpleaños de Lord en 1.982, y en reconocimiento a sus aportaciones, el Educational Testing Service organizó un simposium en el mes de Mayo fruto del cual es a su vez un volumen titulado "*Principals of Modern Psychological Measurement*" editado por Howard Wainer y Samuel Messick (1.983) que contiene, además de un conjunto valioso de contribuciones en torno a algunos de los principales temas de la

obra de Lord (no solo sobre TRI), una lista completa de sus publicaciones hasta esa fecha.

Las contribuciones de Lord a la Psicometría serían más que muy considerables por solo una de sus tres obras principales que han ido marcando etapas principales del desarrollo de teoría psicométrica. Pero además de ellas y de más de un centenar de trabajos, debe reconocerse el papel institucional que ha desempeñado. Por un lado, como presidente de la División de Evaluación y Medida de la APA y de la Sociedad Psicométrica, y también en la formación y crecimiento del "*Psychometric Research Group*" del "*Educational Testing Service*". Por otro, por su capacidad para detectar, rodearse y contribuir a la formación de jóvenes talentos que andados los años se han constituido en principales figuras del campo. Entre ellos, para mencionar sólo unos pocos, debe citarse al sueco Karl Jöreskog, al sudafricano Michael Browne, al danés Erling Andersen, a la japonesa Fumiko Samejima, al alemán Walter Kristof, y al americano Melvin Novick. Otros muchos como Aitken, Birnbaum, Bock, Gleser, Jennrich, Keats, Levine, Meredith, Stcliffe y Zinnes trabajaron también en el importante núcleo de desarrollo que Lord contribuyó a crear en el "*Educational Testing Service*" (Wainer, 1.983). Esta extensa red de 'discípulos' y colaboradores ha tejido el principal desarrollo y aplicación de la teoría de la respuesta al ítem a partir de los trabajos clásicos de Lord.

Jöreskog fue reclutado por Lord en función de sus prometedores trabajos acerca de las bases del análisis factorial. Desde luego Lord no se equivocó con él. Quizás se trate de uno de los autores más leídos y citados en todo el campo de los modelos y las técnicas de análisis multivariadas. A Jöreskog se deben, en buena parte junto a Sörbom, un amplio conjunto de desarrollos de los modelos de medición latentes que implican estructuras factoriales, combinados con los modelos de ecuaciones estructurales. Una combinación que es conocida con la popular denominación de modelos LISREL y que se sitúa actualmente en la cima de los desarrollos de modelos multivariados en ciencias sociales. También ha participado en numerosos trabajos con Werts y con Linn.

Andersen siguió el programa trazado por las aportaciones relativas a la Teoría de la Respuesta al Ítem, concentrando su tarea en los modelos de rasgo latente, denominación inicial de este conjunto teórico. En 1.973 publicó un trabajo sobre tests de bondad de ajuste para los modelos de

Rasch, en 1.977 se ocupó de estadísticos suficientes para los modelos de rasgo latente, y también en ese año, junto a M. Madsen, de la estimación de parámetros de la distribución de población latente.

Samejima también ha permanecido en la línea de desarrollar la teoría de la respuesta al ítem, con trabajos sobre el modelo de ojiva normal (1.973), un método de estimación de las curvas características de ítem usando el estimador de la capacidad de máxima verosimilitud (1.977), y el uso de la función de información en los tests personalizados (1.977). Quizás sea Samejima quien más submodelos y variantes de la TRI ha desarrollado, a decir de revisiones recientes del campo. Una de las razones que se aducen para no clasificar los submodelos de la TRI por el nombre de su creador es que, como sucede especialmente en el caso de Samejima, existirían diversas variantes importantes con un mismo apellido.

Novick, el coautor con Lord de "*Statistical Theories of Mental Test Scores*", ha dado a la luz otros importantes trabajos como "Statistics as Psychometrics" (1.980), "Statistical methods for educational and psychological research", junto a P.H. Jackson (1.974), y "The use of more realistic utility functions in educational applications", con D.V. Lindley (1.978). También ha trabajado en la extensión de la aproximación de tests referidos al criterio, por ejemplo en un artículo de 1.973 "Toward an integration of theory and method for criterion-referenced tests" firmado con Hambleton.

Birnbaum es conocido por su capítulo "*Tests scores, sufficient statistics and the information structure of tests*" en el trabajo de Lord y Novick del año 1.968, cuyas importantes contribuciones ya han sido parcialmente comentadas. Book es también autor de trabajos en el marco de la TRI, y quizás especialmente citado por su revisión en el *Annual Review of Psychology* en 1.971 sobre test mentales. También Keats fue autor, en el año 1967, de la revisión sobre tests mentales en el *Annual Review*; además publicó una introducción a la psicología cuantitativa que ha sido traducida al castellano en 1.971. Por último, Zinnes es sobre todo conocido por su teoría básica de la medida, publicada con Suppes en el *Handbook de Psicología Matemática* de Luce, Bush y Galanter de 1.963. La mayoría de todos estos autores, discípulos y colaboradores de Fred Lord, contribuyeron con algún capítulo a "*Principals of Modern Psychological Measurement*" junto a otros como Robert L. Thorndike, Anne Anastasi, Warren S. Torgerson, Norman Cliff,

Ledyard R. Tucker, Erling B. Andersen, John B. Tukey, y Robert L. Linn, por destacar algunos nombres.

Evidentemente la Teoría de la Respuesta al Ítem ha surgido y se ha desarrollado poderosamente en parte debido a su capacidad para dar respuestas a cuestiones para las que la teoría clásica de tests tendría muy difícil uso. En ese sentido una serie de trabajos destacan en esa extensión y aplicación de la TRI a nuevos problemas. Pueden citarse el libro ya mencionado de Lord en 1.980 y también, antes, el artículo "Item characteristic curve solutions to three intractable testing problems" de Marco de 1.977 y el "Solving measurement problems with the Rasch Model" de Wright de 1.977.

En 1.982 la revista *Applied Psychological Measurement* dedicó un número especial a la teoría de la respuesta al ítem que fue titulado "Advances in Item Response Theory and Applications", teniendo por editores a Ronald K. Hambleton y Wim J. Van der Linden. En el número, además de una introducción general de los editores del mismo, se recogían otros siete trabajos de R.P. McDonald; G.H. Fischer y A.K. Forman; R.J. Mokken y C. Lewis; R.D. Bock y R.J. Mislevy; E.B. Andersen; el editor de la revista, David J. Weiss; y el mismo Frederic M. Lord. La introducción es de Hambleton y Van der Linden, y tiene la virtud de ordenar algunas ideas generales acerca de la teoría de la respuesta al ítem y su relación con la teoría clásica de los tests mentales.

La teoría de la respuesta al ítem ha sido desarrollada originalmente, en su mayor parte, orientada hacia ítems dicotómicamente valorados. Típicamente ítems con respuesta verdadera tales como los propios de un test de inteligencia, un test de aptitud o una prueba de rendimiento. A ese nivel puede encontrarse formulada en el principal volumen de Lord y Novick de 1.968. Posteriormente ha sido extendida para ítems politómicamente valorados y con otros formatos en los trabajos de Samejima de 1.969 y de 1.972, y de Bock de 1.962 y de 1.972, a los que seguirán muchos otros. También se ha ocupado Samejima del tratamiento de la teoría de la respuesta al ítem considerando un nivel de respuesta en términos de variable continua. Además, se han ofrecido una serie de extensiones de la teoría desde un punto de vista multivariado en los trabajos de Samejima de 1.974, de Sympson de 1.977 y de Whiteley de 1.980.

Uno de los autores que más ha trabajado en efectuar revisiones y presentaciones generales de la Teoría de la Respuesta al Ítem es R.K. Hambleton. En esta línea ha publicado trabajos en 1.977, 1.978, 1.979 y en 1.983. También se ha interesado Hambleton, entre otras muchas facetas de la TRI, por la relación entre la TRI y la medición referida al criterio, primero en un artículo de 1.973 firmado con Novick, y luego en un artículo publicado en solitario diez años después en *Applied Psychological Measurement*.

En 1.985 Ross E. Traub y Y. Raymond Lam realizan una revisión de la literatura que titulan "*Latent structure and ítem sampling models for testing*", en ella incluyen la teoría de la respuesta al ítem y el análisis de clase latente como las dos grandes ramas del análisis de estructura latente, que consideran junto a los modelos de muestreo de ítems. Como primera constatación Traub y Lam señalan que el interés de la investigación se ha desplazado desde la cuestión de la fiabilidad, que ocupaba el lugar central cinco o más décadas atrás, hacia la temática a la que dedican su revisión. Particularmente observan que en la última década la investigación se ha centrado, con un considerable incremento, en la teoría de la respuesta al ítem y los modelos de muestreo de ítems, como ellos los llaman, sus condiciones y sus aplicaciones. Sin embargo, la mayoría de los modelos desarrollados han sido pensados para tests compuestos de ítems dicotómicamente puntuados.

Hay disponibles un número de introducciones y de presentaciones sumarias de la TRI clásica. Birnbaum (1.968), Fischer (1.974), Wright y Masters (1.982) y Hambleton y Swaminathan (1.985) presentan introducciones a los varios modelos de teoría de la respuesta al ítem. Thissen y Steinberg (1.986) presentan una taxonomía de esos modelos. También pueden contarse las exposiciones de Santisteban (1.984; 1.990), Linn (1.988), Lord y Stocking (1.988), Douglas (1.988) y Wright (1.988).

Con la mención de algunas de las presentaciones generales de la TRI concluimos esta breve aproximación histórica. La TRI es probablemente el campo de trabajo de más intenso desarrollo en segunda mitad del siglo XX. Aunque sus raíces se sitúan hacia mediados del siglo, es durante las últimas décadas del siglo XX cuando se ha producido una verdadera explosión en trabajos de teoría y aplicación de la teoría de la respuesta al ítem, con una fuerte diversificación abarcando decenas de nuevos

modelos unidimensionales y multidimensionales. Una serie de empresas norteamericanas dedicadas a la elaboración y comercialización de tests, además claro de organismos como el E.T.S. (*Educational Testing Service*) desde mucho antes, están utilizando ya TRI como procedimiento común de trabajo en la construcción de bancos de ítems. No es una predicción arriesgada considerar con Carroll (1.988) que el futuro de la teoría y la aplicación psicométrica durante los próximos años estará ligado al desarrollo de este cuerpo teórico. Sin embargo, la implantación de la TRI en el trabajo diagnóstico y métrico de la psicología ha sido relativamente escasa -comparado con el uso de la Teoría Clásica de Tests- durante las últimas décadas del siglo XX.

Dado que la teoría de la respuesta al ítem supone un rasgo latente, subyacente y continuo, de carácter inobservable, la teoría de la respuesta al ítem puede clasificarse como una de las dos grandes ramas de los modelos de rasgo latente, también denominados modelos de estructura latente o análisis de la estructura latente.

3. Concepto y tipos principales de modelos de estructura latente.

Se ha denominado análisis de la estructura latente, o modelos de rasgo latente, a un conjunto particular de teoría psicométrica orientado a tratar con puntuaciones de tests de aptitudes bajo dos supuestos: el de la existencia de uno o más rasgos latentes y el de la independencia estadística de las respuestas de los sujetos a los n ítems.

El conjunto de respuestas que un sujeto da a los ítems de un test se denomina patrón de respuestas. Hay 2^n patrones de respuestas diferentes posibles para un test con n ítems dicotómicamente valorados. "El supuesto básico del análisis de estructura latente es que cada sujeto en una población puede caracterizarse por su puntuación en una o más variables latentes inobservadas, tal que *en una subpoblación de sujetos, cada uno con la misma puntuación en cada variable latente*, las repuestas a los ítems en el test son mutuamente estadísticamente independientes. (Traub y Lam, 1.985). Esta formulación verbal del *supuesto de independencia local* puede expresarse matemáticamente:

$$P(u|\theta) = P(u_1|\theta) \cdot P(u_2|\theta) \cdot \dots \cdot P(u_n|\theta)$$

representando u el patrón de respuesta de un sujeto, u_i es la puntuación dicotómica al ítem i , y θ es el vector de puntuaciones del sujeto en las variables latentes, $P(u|\theta)$ es la probabilidad condicional del patrón de respuesta, y $P(u_i|\theta)$ es la probabilidad de la respuesta, cero o uno, al ítem i dadas las puntuaciones del examinado en las variables latentes. De este modo la ecuación representa la probabilidad de un patrón de respuesta en función de las probabilidades de respuesta a los ítems individuales del test, poniendo de manifiesto que las probabilidades de respuesta a los ítems son independientes entre sí dado un nivel en el rasgo o vector de rasgos latentes.

Según Traub y Lam (1.985), seguiremos la convención de denominar Análisis de Estructura Latente (AEL) al conjunto de modelos más amplio que incluye a la teoría de la respuesta al ítem (TRI) y el análisis de clase latente (ACL). De acuerdo con Traub y Lam (1.985) "Puede trazarse una distinción entre dos tipos de modelos de estructura latente, los de la teoría de la respuesta al ítem (TRI) y los del análisis de clase latente (ACL). La base de la distinción reside en la naturaleza supuesta de la distribución de θ en la población de examinados. Esta distribución se toma por continua para la TRI y por discreta para el ACL. De este modo, la TRI es considerada apropiada para los tests psicológicos de capacidad o actitud, en que los examinados de una población son vistos como continuamente distribuidos sobre las variables latentes subyacentes. El análisis de clase latente, por otra parte, es apropiado para aquellas situaciones en que los examinados son vistos como pertenecientes a solo dos, o por lo menos muy pocos grupos distintos. Este puede ser el caso, por ejemplo, en tests de maestría (mastery)." Esta distinción ha sido reconocida implícitamente por Lord y Stocking (1.988) al afirmar que la "teoría de la respuesta al ítem cae dentro de la clase general de modelos de rasgo latente".

A pesar de la lógica y ordenada clasificación de Traub y Lam (1.985), la denominación más general de todo este conjunto teórico desde el trabajo de Lazarsfeld de 1.950 fue durante mucho tiempo la de modelos o teoría del rasgo latente, y, desde la publicación del libro de Lord de 1.980 "*Applications of item response theory to practical testing problems*", esa denominación tendió a sustituirse por la de teoría de la respuesta al ítem. Sin

embargo debería mantenerse la denominación de 'modelos de rasgo latente' para el caso general y distinguir dentro de ellos el 'análisis de clase latente', cuando se trata de θ discreto, y la 'teoría de la respuesta al ítem', cuando se trata de θ continuo. La 'teoría de la respuesta al ítem' se refiere al conjunto de modelos que ponen en relación un rasgo latente continuo con la probabilidad de respuesta correcta a un ítem, manteniendo el supuesto de independencia local y describiendo esa relación por una función generalmente logística, por lo general de uno, dos o tres parámetros.

En la *figura 9* (ver Capítulo 4) se ha presentado un esquema ordenador de una selección de teorías y modelos enmarcables dentro del análisis de la estructura latente. En la figura puede apreciarse como los modelos de rasgo latente hacen dos supuestos (la existencia de θ y la independencia estadística local de la respuesta a los ítems) y como estos modelos se subdividen en dos. Por un lado el 'Análisis de Clase Latente' cuando θ se supone discreta; por otro, la '*Teoría de la Respuesta al Ítem*', cuando θ se supone continuo. A su vez dentro de la TRI se distinguen dos clases de modelos en función de que trate de θ unidimensional o multidimensional (es decir, uno o más rasgos latentes continuos subyacentes). Dentro de cada una de estas categorías todavía es posible distinguir los modelos diseñados para ítems dicotómicos o politómicos -y otras variantes-. Sin embargo, a pesar de estas distinciones y otras más que efectúa la figura 8, la mayoría del trabajo en TRI clásica se ha desarrollado con modelos unidimensionales para ítems dicotómicos y con funciones logísticas de uno o tres parámetros, campo al que restringiremos la discusión en la presentación de algunos conceptos básicos.

4. Algunos conceptos básicos de la Teoría de la Respuesta al Ítem.

Como un postulado central, la TRI considera que la probabilidad de la respuesta correcta a un ítem puede representarse adecuadamente por una función matemática, usualmente por una función logística. Esta descripción matemática de la relación entre el nivel de una persona en un rasgo latente y su desempeño ante un ítem se denomina función de respuesta al ítem, curva característica del ítem o línea huella (Lord y Stocking, 1.988). Quizás sería más preciso reservar la denominación

"función" para la expresión algebraica de esa relación y reservar la expresión "curva característica del ítem" (cci) para la forma gráfica de esa función, sin embargo en los trabajos de TRI no ha sido infrecuente un uso genérico de ambas expresiones.

Esa función matemática logística depende -en los modelos clásicos más importantes- de una serie de parámetros del ítem, -de uno a tres,- y de un solo parámetro θ de la persona. El parámetro θ de la persona se refiere a su capacidad, a su aptitud, en general, a su ubicación en la dimensión latente θ . Los parámetros del ítem constituyen las características psicométricas principales de los mismos, y ubican también al ítem en el continuo θ de la dimensión que mide.

Función logística de tres parámetros. En términos de Lord (1.980) puede expresarse como:

$$P \equiv P(\theta) = c + \frac{1 - c}{1 + e^{-1.7a(\theta - b)}}$$

"donde a , b y c son parámetros que caracterizan el ítem, y e es la constante matemática 2.71828...[...]. El parámetro c es la probabilidad de que una persona completamente carente de habilidad ($\theta = -\infty$) responda el ítem correctamente. Se lo denomina parámetro de adivinación, de conjetura, o nivel de puntuación de pseudo-suerte. Si un ítem no puede ser contestado correctamente por adivinación, entonces $c = 0$.

El parámetro b es un parámetro de localización: Determina la posición de la curva a lo largo de la escala de capacidad [a lo largo del continuo de θ]. Se lo denomina dificultad del ítem. Cuanto más difícil el ítem, más la curva se sitúa hacia la derecha [hacia los valores altos de θ]. La curva logística tiene su punto de inflexión en $\theta = b$. Cuando no hay adivinación b es el nivel de capacidad donde la probabilidad de una respuesta correcta es 0'5. Cuando hay adivinación, b es el nivel de capacidad donde la probabilidad de una respuesta correcta está a medio camino entre c y 1.

El parámetro a es proporcional a la inclinación de la curva en el punto de inflexión {esta inclinación realmente es $0.425a(1 - c)$ }. Así el parámetro a representa la potencia discriminativa del ítem, el grado en que la respuesta al ítem varía con el nivel de capacidad." (Lord, 1.980; pag. 12-13).

En la *figura 13* (tomada de Lord, 1.980) se muestran las curvas características de los ítems de un test, y en la *figura 14* (también debida a Lord, 1.980) se representa de modo gráfico el significado de los parámetros de los ítems.

Figuras 13 y 14 debidas a Lord (1980).

Figura 13.

Curvas características de ítem de los ítems de un test de aptitudes.
(Lord, 1.980)

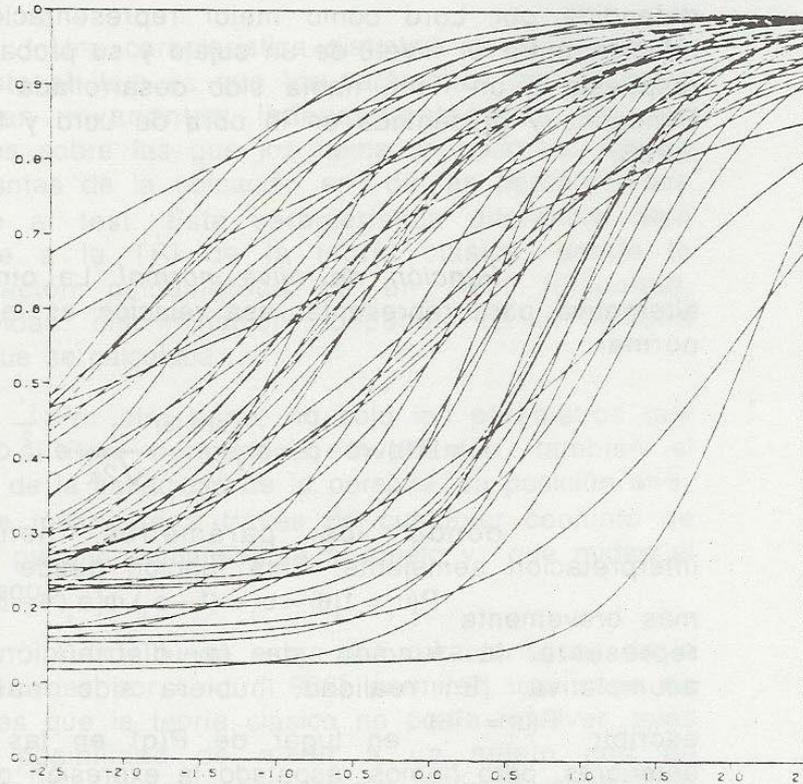
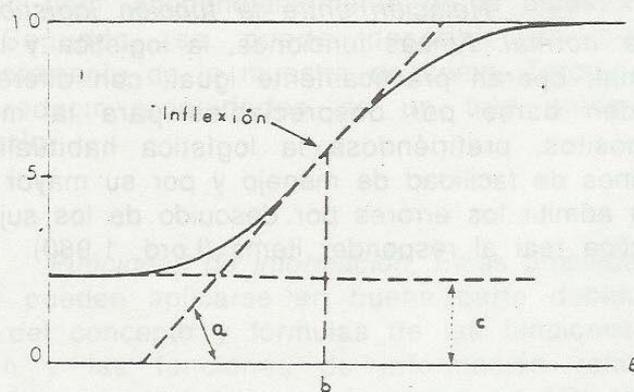


Figura 14.

Representación del significado de los parámetros de un ítem sobre una curva característica de ítem. (Lord, 1.980)



Esta función logística de tres parámetros defendida por Lord como mejor representación de la relación entre el nivel θ de un sujeto y su probabilidad de respuesta a un ítem, había sido desarrollada por Allan Birnbaum, y presentada en la obra de Lord y Novick de 1.968.

Función de ojiva normal. La otra función alternativa clásica para representar esa relación es la de ojiva normal

$$P \equiv P(\theta) = c + (1 - c) \int_{-\infty}^{a(\theta-b)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

donde los parámetros tienen una interpretación semejante.

Esta función puede escribirse más brevemente:

$$P(u_i = 1|\theta) = c_i + (1 - c_i) \Phi[a_i(\theta - b_i)]$$

si Φ representa la función de la distribución normal acumulativa.

(En realidad, hubiera sido más preciso escribir

$$P(u_i = 1|\theta)$$

en lugar de $P(q)$ en las fórmulas anteriores, pero hemos respetado la expresión original).

Relación entre la función logística y la de ojiva normal. Ambas funciones, la logística y la de ojiva normal, operan prácticamente igual, con diferencias que pueden darse por despreciables para la mayoría de propósitos, prefiriéndose la logística habitualmente por razones de facilidad de manejo y por su mayor capacidad para admitir los errores por descuido de los sujetos en la práctica real al responder ítems (Lord, 1.980).

Calibración de los ítems e independencia de muestra. Al proceso de obtención de los parámetros de un ítem se lo conoce como calibración del ítem.

Una característica distintiva de la teoría de la respuesta al ítem es que los parámetros de los ítems permanecen invariantes independientemente de las muestras sobre las que los ítems se apliquen, siendo independientes de la ubicación en θ de las personas que responden al test. Esta característica diferencia bien claramente a la TRI de la teoría clásica, donde la caracterización psicométrica de un ítem (dificultad, homogeneidad, discriminación), dependía de la muestra sobre la que se calculaba.

Por otra parte, no solo los parámetros que describen al ítem permanecen invariantes, también el parámetro de la capacidad de la persona, su posición en θ , permanece invariante a través de cualquier conjunto de ítems que puedan administrarse al sujeto y que midan el mismo rasgo θ .

Desde luego estas dos características de la TRI, enfatizadas por Linn (1.988) permiten una serie de aplicaciones que la teoría clásica no podía resolver, tales como la posibilidad de medir a un sujeto con un subconjunto "a la medida" de ítems (tailored testing) de un banco de ítems. Lord y Stocking (1.988) han sintetizado las tres ventajas principales de la TRI. Primero, se puede describir un ítem independientemente de la muestra de sujetos. Segundo, se puede describir un sujeto independientemente de la muestra de ítems. Tercero, se pueden predecir propiedades de un test antes de administrarlo.

Funciones de información. Estas propiedades de la TRI pueden aplicarse en buena parte debido al desarrollo del concepto y fórmulas de las funciones de información y las funciones de información relativa. Aunque la idea seminal de estas funciones en TRI puede encontrarse ya en la tesis de Lord de 1.951, un desarrollo sistemático de las mismas aparece en Birnbaum (1.968). Hay funciones de información de los ítems y funciones de información de los tests.

Una *función de información del ítem* se define como

$$I(\theta, u_i) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

donde:

$P_i(\theta)$ es la función de respuesta al ítem,

$$Q_i(\theta) \equiv 1 - P_i(\theta),$$

y,

$P'_i(\theta)$ es la derivada con respecto a θ de la función de respuesta al ítem.

La *función de información del test* se define como

$$I(\theta) = \sum_{i=1}^n I(\theta, u_i)$$

siendo n el número de ítems en el test.

Sin entrar en otras interpretaciones, las funciones de información de un ítem o de un tests proveen información sobre la efectividad del ítem o del test en cada nivel del rasgo que está siendo medido. La función de información del test puede interpretarse como la máxima información disponible de un test, independientemente del procedimiento de puntuación que se utilice.

De acuerdo con la anterior definición de $I(\theta)$, cada ítem contribuye a la efectividad del test independientemente de las contribuciones de los demás ítems.

Función de eficiencia relativa. Por último, dados dos tests (o partes de tests, o ítems de tests) que miden el mismo rasgo θ , la función de relativa eficiencia del test y frente al x se define por

$$R(x, y; \theta) = \frac{I(\theta, y)}{I(\theta, x)}$$

Si $R(x, y; \theta) > 1$, entonces el test y da más información que el test x al nivel θ considerado. Las funciones de

eficiencia relativa tienen la virtud de permanecer invariantes frente a cambios monotónicos del modo de medición de θ lo que las convierte en un instrumento muy adecuado para la comparación de tests.

Funciones logísticas de uno y de dos parámetros.

Por supuesto las funciones logística y de ojiva normal de tres parámetros no han sido las únicas funciones de respuesta al ítem que han sido propuestas.

Rasch en 1.960 dio lugar a una familia de modelos mucho más sencillos donde la probabilidad de respuesta correcta de un sujeto a un ítem depende de solo dos cosas: la dificultad del ítem y la capacidad de la persona. Es decir de la ubicación del ítem en θ y de la ubicación de la persona en θ . Este tipo de modelos contienen un solo parámetro para describir al ítem, el parámetro dificultad, y han sido descritos como logísticos de un parámetro (siendo el parámetro $a = \text{cte.}$ y el parámetro $c = 0$). Los modelos de Rasch poseen un conjunto de propiedades psicométricas particularmente valiosas y favorables que los distinguen de otros tipos de modelos de TRI.

Por su parte, Prentice en 1.976 consideró los modelos logísticos y de ojiva normal de dos parámetros, incluyendo en las funciones anteriores $a=1$, $b=0$ y $c=0$, si bien el trabajo con modelos de dos parámetros se ha fundado en el caso $c=0$.

Campos de aplicación. Originalmente la TRI nació asociada a la medición de *aptitudes*, (capacidad verbal, inteligencia, etc.) sin embargo, muy pronto se ensayaron aplicaciones en el campo de las *actitudes* (Lazarsfeld, 1.959) en el que se ha aplicado a constructos como el de satisfacción laboral (Parsons, 1.979).

La TRI ha destacado por su utilidad para determinados propósitos tales como evaluar la fidelidad psicométrica de una versión traducida de un test (Hulin, Drasgow y Komocar, 1.982), los ya mencionados tests de "maestría", la "igualación" ("equating") de tests -proceso y resultado de establecer las equivalencias entre las puntuaciones de dos tests sobre un mismo contenido-, la estimación de distribuciones de capacidad y

otras. Pueden destacarse sus aplicaciones en el estudio de la "adecuación de la medición", los "sesgos de ítems y de tests", y los llamados "tests a la medida".

Tests e ítems sesgados. Un test o un ítem están sesgados contra una subpoblación determinada cuando los resultados de esa subpoblación difieren negativamente respecto a otra tomada como referencia. Por ejemplo, en algunos tests es posible encontrar que los promedios de un grupo racial son consistentemente menores que los de otro, o bien que los resultados de un sexo son inferiores a los de otro. La constatación de esas diferencias reiteradamente es una cuestión de hecho, la interpretación de las mismas en términos de diferencias en capacidades u oportunidades y las consecuencias sociales que pueden derivarse exceden el campo psicométrico y alcanzan cuestiones éticas y políticas. Desde el punto de vista de la Teoría de la Respuesta al Ítem, "un ítem da medidas equivalentes a través de subpoblaciones si la relación entre la probabilidad de una respuesta correcta y la capacidad es virtualmente idéntica en todas las subpoblaciones" (Hulin, Drasgow y Parsons, 1.983).

Adecuación de la medida ("Appropriateness measurement"). "Aún cuando un test particular no contenga ítems sesgados contra ninguna subpoblación identificable, y ofrezca una medición excelente para todas las muestras de examinados, una puntuación de este test puede ser seriamente deficiente como estimador de la capacidad de un pequeño número de sujetos atípicos" (Hulin, Drasgow y Parsons, 1.983). Entre esos sujetos se encuentran por ejemplo personas extraordinariamente capaces, inteligentes o creativas que pueden enfrentar algunas o todas las tareas de un test a un nivel y de un modo muy distinto al de las demás personas, induciéndoles a resultados que no reflejan su ubicación en la dimensión medida. Otros ejemplos incluyen personas descuidadas o particularmente inhábiles en alguna habilidad secundaria al test, personas que confunden el mecanismo de respuesta y ubican mal las respuestas, personas que por algún motivo conocen previamente la prueba o una porción de la misma cuando esto no debe suceder o personas que utilizan medios no honestos para elaborar sus respuestas.

Tests a la medida (Adaptive o tailored testing). La idea esencial de los tests a la medida supone aplicar a cada sujeto solo aquellos ítems que resultan adecuados a ese sujeto. Aquellos ítems que están muy por debajo de la capacidad de un sujeto y aquellos que están muy por encima de su capacidad no aportan información sobre la ubicación del sujeto en el continuo de capacidad. Esos ítems pueden resultar completamente superfluos y suponen desperdiciar tiempo y esfuerzo. Por el contrario, un conjunto de ítems graduados en el entorno del nivel del sujeto resultan útiles al permitir ubicarlo adecuadamente. Este objetivo de utilizar solo los ítems adecuados al sujeto se pretende conseguir operativamente utilizando la información recogida en un ítem para escoger el nivel de dificultad del siguiente. Supongamos que a un sujeto se le presenta un ítem d con un nivel de dificultad dado. Si el sujeto responde adecuadamente el ítem d entonces se le ofrece un ítem f de mayor dificultad. Si el sujeto supera f entonces se le presenta otro de mayor dificultad, pero si falla f se selecciona el tercer ítem en un nivel de dificultad entre el del ítem d y el del ítem f. Si desde el principio el sujeto falla el ítem f entonces se le presenta un ítem de menor dificultad que f. En cada momento se aprovecha toda la información disponible para conjeturar la ubicación del sujeto en el continuo de capacidad y presentarle los ítems más adecuados. El lector puede pensar que este objetivo y esta dinámica bien podría efectuarse bajo teoría clásica de tests, sin embargo, la dificultad estriba en que la teoría clásica no puede colocar a sujetos medidos con distintos subconjuntos de ítems en una misma escala. En todo caso la descripción anterior es únicamente una simplificación del modo en que opera la TRI en la selección de ítems en cada paso utilizando las funciones de información.

Relación de la TRI con la validez. La TRI no es una teoría sobre la validez de los tests, ni tampoco una teoría sobre su fiabilidad. La TRI es un cuerpo teórico relativo al modo en que la puntuación en los tests puede ser escalada independientemente de una muestra de sujetos o de ítems concretos. Según Hulin y colaboradores "Las teorías de la medida deben ser diferenciadas de las teorías sustantivas o de contenido que especifican o definen los significados de los rasgos subyacentes en términos de sus relaciones con otras variables y conductas. Así la TRI puede ser usada para estimar la posición de un individuo en el rasgo latente θ pero no puede orientar la validación de constructo del rasgo" (Hulin, Drasgow y Parsons, 1.983). Esto no significa que para los tests

elaborados desde el punto de vista de la TRI pueda prescindirse del proceso de validación de los mismos. Los tests son útiles sólo para los propósitos para los que han *demostrado* mediante investigación empírica ser útiles, es decir sólo en aquellos contextos y propósitos en que han sido validados.

Con este énfasis en la necesidad de validación de los tests cerramos una primera aproximación a algunos de los aspectos de ubicación, perspectiva histórica y conceptos básicos de la teoría de la respuesta al ítem. El lector deberá acudir a las obras originales citadas en el texto para una introducción explícita y suficiente a la misma. Por nuestra parte esperamos haber puesto de manifiesto el lugar central de la teoría de la respuesta al ítem en la estructura global del campo psicométrico, su relevancia y actualidad como campo de investigación, algunos de sus principales desarrollos históricos y una primera enunciación de algunos de sus conceptos clave.