

ESTRATEGIAS Y TÉCNICAS CUANTITATIVAS PARA EL ESTUDIO DEL DOMINIO AFECTIVO EN MATEMÁTICAS

Strategies and quantitative techniques for the study of the affective domain in mathematics

Palacios-Picos, A.

Departamento de Psicología de la Universidad de Valladolid, Facultad de Educación de Segovia

Resumen

La medida del dominio afectivo matemático supone un aspecto de vital importancia en la investigación matemática. No obstante, no ha tenido el suficiente desarrollo en la formación de los futuros investigadores en este campo de las matemáticas. Nuestro objetivo en las líneas que siguen es presentar los diferentes modelos de medida y las teorías que los sustentan comenzando con la Teoría Clásica de Medida (TCM). Esta teoría está siendo sustituida por nuevos modelos como la Teoría de Respuesta al Ítem (TRI) o por aquellos derivados de las ecuaciones estructurales o modelos SEM, que se presentan en sus aspectos fundamentales recalcando sus principios, sus técnicas y sus aportaciones. Para finalizar, se resumen las propuestas de medida más relevantes relacionadas con las actitudes hacia las matemáticas y la ansiedad matemática con especial interés en las basadas en estas teorías de la TRI o los modelos de ecuaciones estructurales (SEM).

Palabras clave: Modelos de ecuaciones estructurales, Teoría de Respuesta al Ítem, medida, dominio afectivo matemático, actitudes hacia las matemáticas, ansiedad matemática.

Abstract

The mathematical measure of affective domain is a vital aspect in mathematical research. However, it has been a field of study that has not been sufficiently developed nor has taken the place it should in the training of future researchers. Our main goal is to present the different measurement models and theories that support beginning with the Classical Test Theory (CTT). This theory is being overtaken by new models like the Item Response Theory (IRT) or those derived from or Structural Equation Models (SEM). They are presented in summarized through its principles, techniques and contributions. Finally, summarizes focused on the two most significant aspects of mathematical affective domain such as attitudes towards mathematics and mathematics anxiety proposals as most important in recent years with special emphasis on those based on these latest theories IRT or Structural Equation Modeling.

Keywords: Structural equation models, item response theory, measurement, mathematical affective domain, attitudes towards mathematics, math anxiety.

MODELOS CUANTITATIVOS DE MEDIDA EN CIENCIAS SOCIALES

Teoría clásica de la medida

Los trabajos pioneros de Spearman de principios de siglo pasado (Spearman 1904, 1907, 1913) supusieron el inicio de lo que hoy conocemos como la Teoría Clásica de los instrumentos de medida. Venían a solucionar un problema de los test psicológicos que por entonces comenzaron a desarrollarse con no poco éxito: determinar el grado de exactitud de sus medidas y asegurar el mayor grado de precisión posible. La idea de Spearman no pudo ser más sencilla ni más cercana al sentido común: la puntuación de un sujeto (su puntuación empírica) está formada por la puntuación verdadera y el error que cometemos en esta medida.

Dado que no podemos conocer directamente ambos componentes, es necesaria su estimación a partir de un conjunto de supuestos; señalamos los más importantes. Uno de ellos define la puntuación verdadera como la esperanza matemática de las posibles puntuaciones empíricas. Es decir, pasemos infinitas veces el mismo test, al mismo sujeto, calculemos su valor medio y habremos obtenido la puntuación verdadera. Un segundo supuesto mantiene que no existe relación entre el tamaño del error que cometemos y la cuantía de la puntuación verdadera; es decir, puede haber tanto puntuaciones verdaderas muy elevadas y errores muy pequeños como puntuaciones muy pequeñas con errores elevados.

Como señala Muñiz (2010), estos supuestos junto con el modelo lineal que se deriva de ellos y la definición de los test paralelos constituyen el núcleo principal de la teoría clásica de medida. La importancia del instrumento frente a otros elementos de la medida determinó el interés de los investigadores por el cálculo de índices de consistencia de las pruebas así como por la importancia de los grupos normativos. Desde esta perspectiva clásica se considera un instrumento fiable si existe concordancia entre las puntuaciones de un mismo sujeto en diferentes momentos. Se suele recurrir al método del test-retest como procedimiento de cálculo en el que los mismos sujetos son medidos en dos o más ocasiones diferentes, siendo la correlación de Pearson la cuantificación de su fiabilidad. No obstante, ante la dificultad que plantea este cálculo, se ha tendido a sustituir el cálculo de la fiabilidad como el grado de repetición de una medida por la homogeneidad de las preguntas que componen la escala. Para el cálculo de la fiabilidad en este segundo caso, se recurre a técnicas como de las dos mitades o a la correlación de los ítems entre sí, siendo en este último caso el alfa de Cronbach el estadístico más utilizado (con todos los inconvenientes que conlleva).

Aunque la validez de la medida ocupó un espacio menos relevante que el concepto de fiabilidad dentro de esta Teoría Clásica, no deja de ser una característica buscada por cualquier investigador. La validez de una escala está determinada por el grado en el que se mide realmente lo que se dice medir. Tras la simpleza de esta definición, se esconde una dificultad, en muchos casos insalvable, pues no siempre es fácil definir nuestro objeto de estudio. Para operativizar esta característica se utilizan diferentes medidas como son la validez de contenido, la validez de criterio o la validez de constructo (Tabla 1).

Tabla 1.- Diferentes tipos de validez

Validez de contenido	Grado en el que el instrumento de medida presentada una muestra adecuada de ítem referidos a los contenidos que se pretende medir. Se basa fundamentalmente en el juicio de expertos que definen el grado en el que las preguntas representan los contenidos a evaluar. Aunque no suelen utilizarse métodos cuantitativos para su cuantificación, existes índices para esta tarea como el Índice de
----------------------	---

Validez de Contenidos de Lawshe (1975)

Validez de criterio	Se define como la concomitancia que muestra la escala con variables ajenas a la prueba (criterio), que tomamos como referencia de validez. Generalmente, dicho criterio ha sido testado previamente en cuanto su validez con respecto a lo que queremos medir.
Validez de constructo	Dado que la mayor parte de nuestros instrumentos de medida se refieren a aspectos que carecen de realidad (constructos), para establecer su validez se deben evaluar las predicciones e hipótesis que se derivan del mismo y de la teoría que lo sustenta. En este sentido, la validez de constructo respondería a la pregunta ¿la medida se ajusta a la teoría sobre la que se construyó la escala y a sus posibles predicciones? Como concepto general de validez, es el más complejo de calcular y englobaría a algunos de los otros tipos.

La fiabilidad y la validez dependen, en gran medida, de la correcta elección de los ítems que forman la escala, lo que obliga a un análisis previo de estos elementos. Uno de los objetivos fundamentales es conseguir preguntas relevantes, claras y discriminantes evitando, además, ciertos sesgos de respuestas (sobre todo en las escalas de actitudes), entre los que señalamos la *acquiescencia* y la *deseabilidad social*. En el primer caso, se trata de disminuir o eliminar la tendencia a mostrar acuerdo con cualquier pregunta o incluso a puntuar las opciones intermedias de las escalas (ausencia de compromiso con lo afirmado). Este sesgo se relaciona en muchos casos con la falta de claridad de las preguntas o con su mala redacción. El concepto de *deseabilidad social* hace referencia a la tendencia a contestar no en función de lo que uno piensa sino sobre lo que se considera más aceptable socialmente.

Teoría de Respuesta al Ítem

En las últimas décadas hemos asistido a un cambio en el modo de abordar el estudio psicométrico de los instrumentos de medición en todas las ciencias sociales. Estos nuevos procedimientos no conllevan incompatibilidad sino coexistencia con los planteamientos de la Teoría Clásica de medida, permitiendo mejorar la consistencia de las escalas de medida (Tabla 2).

Tabla 2. Diferencias entre la Teoría Clásica de los Test y la Teoría de Respuesta al Ítem (Muñiz, 2010)

Aspectos	Teoría Clásica	Teoría de Respuesta al Ítem
Modelo	Líneal	No lineal
Asunciones	Débiles (fáciles de cumplir por los datos)	Fuertes (difíciles de cumplir por los datos)
Invarianza de las mediciones	No	Si
Invarianza de las propiedades del test	No	Si
Escala de las puntuaciones	Entre 0 y la puntuación máxima en el test	Entre $-\infty$ y $+\infty$
Énfasis	Test	Ítem
Relación ítem-test	Sin especificar	Curva característica del ítem
Descripción de los ítems	Índices de dificultad y discriminación	Parámetros a (dificultad), b (discriminación) y c

Errores de medida	Erros típico de la medida común para toda la muestra	(adivinación) Función de información (varía según el nivel de aptitud o habilidad)
Tamaño muestral	Puede funcionar bien con muestras entre 200 y 500 sujetos aproximadamente	Se recomienda más de 500 sujetos, aunque depende del modelo.

De entre estos nuevos procedimientos, es seguramente la Teoría de Respuesta al Ítem (TRI) la de mayor calado y la que posee mayor reconocimiento (Martínez-Arias, 1995). Como en otros muchos casos, surgen como intento de solución de los problemas de la Teoría Clásica de la medida para quien la puntuación final de un instrumento es el resultado de sumar la puntuación real que el individuo tendría en la prueba más el error que cometemos. Como el error de medida se distribuye de manera aleatoria, con un valor medio de cero, cabe suponer que la puntuación observada coincidirá con la verdadera tras pasar la prueba al mismo sujeto infinitas veces y calcular la media final de todas las tomas. Para Attorresi, Lozzia, Abal, Galibert y Aguerri (2009) este planteamiento conlleva un conjunto de presunciones erróneas. La primera es suponer que el resultado al medir una variable es inherente al cuestionario; esto es tanto como decir que el peso de un objeto depende de la balanza que se utilice para pesar. La segunda es asumir que, tanto el cuestionario de medida como las propiedades de las preguntas, están determinadas por las características de los sujetos a los que hemos pasado la escala. Como sugieren Attorresi et al. (2009) sería como pensar que un kilo de acero puede pesar distinto a un kilo de plumas.

La TRI, en contraposición, es un intento de realizar mediciones que no dependan del instrumento de medida; es decir, que sean invariantes con respecto o de las escalas empleadas (Muñiz, 1997). Además, se trataría de disponer de instrumentos de medida cuyas propiedades finales no dependan del grupo normativo (que sea invariante respecto a las personas). Si dos sujetos presentan el mismo nivel de desempeño en un rasgo, tendrán la misma probabilidad de dar la misma respuesta independientemente de la población de referencia. Aspectos ambos que parecen hacer innecesario disponer de un grupo de referencia, evitando el uso de grupos normativos.

No obstante, las fortalezas que manifiesta esta teoría se fundamentan en un conjunto de supuestos estrictos a los que gran parte de los datos empíricos difícilmente se ajustan. Como señalan Hambleton, Swaminathan y Rogers (1991), el resultado final de un sujeto dependerá de la fiabilidad de la escala, que determinará la probabilidad de que el sujeto acierte un ítem determinado. Es decir, cuanto mejor sea la actitud hacia las matemáticas de un estudiante, mayor probabilidad tendrá de puntuar alto en una pregunta relacionada con esta actitud. Esto explica el porqué del nombre de la teoría pues evalúa el ítem y no la prueba en su totalidad. Esta relación entre la habilidad, el conocimiento o la actitud y la probabilidad de acertar el ítem queda plasmada en el concepto central de esta teoría como es la *Curva Característica del Ítem* (Figura 1).

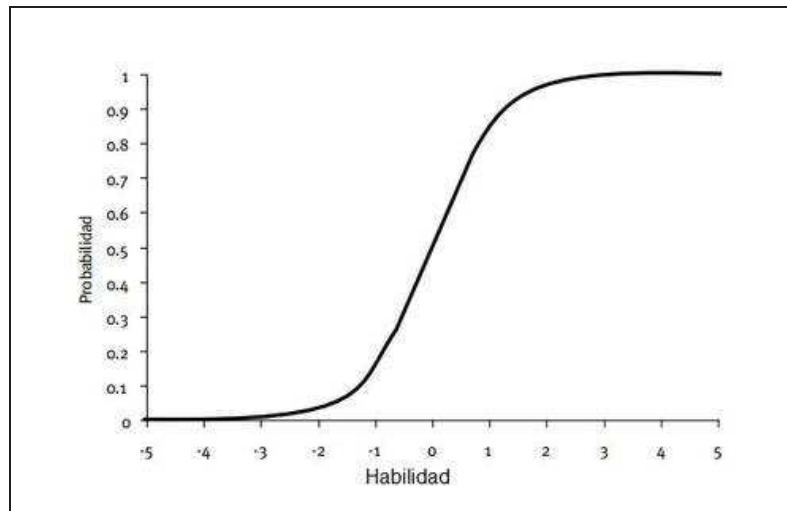


Figura 1.- Curva Característica del Ítem

Esta curva vincula la probabilidad de dar una respuesta (un ítem de la escala de actitudes hacia las matemáticas, pongamos por caso), con el nivel del rasgo latente que se está midiendo (la verdadera actitud del entrevistado). Concretamente, cuanto mejores sean las actitudes hacia las matemáticas, mayor probabilidad de puntuar alto en el ítem. Sin embargo, como señala Hamblenton et al. (1991) este supuesto se fundamenta en la unidimensionalidad de lo que se pretende medir. Es una única *habilidad* (nivel de rasgo) la que determina el resultado del sujeto. Sin embargo, esto no es siempre posible pues, en ciencias sociales, la mayor parte de los constructos dependen de varios factores. No obstante, se han desarrollado modelos no basados en la existencia de un único factor (multidimensionales), pero como señala Attorresi et al. (2009) estos desarrollos son relativamente incipientes y su puesta en práctica resultan costosos y complejos.

Muy relacionado con el supuesto anterior es la necesidad de que la medida de un ítem debe ser independiente de lo que mida otro de la misma escala (*independencia local*). Esto supone simplemente que el desempeño de un ítem debe ser exclusivamente la habilidad, el conocimiento o la opinión del encuestado. Como hemos indicado, el supuesto de *unidimensionalidad* y la *independencia local* son aspectos complementarios, dado que, sí dos ítem no son independientes, debemos asumir que otro factor, que no es el que se pretende medir, está determinando el nivel de desempeño, violándose el supuesto de unidimensionalidad (Lord y Novick, 1968).

Modelos de medida relacionados con las ecuaciones estructurales

Actualmente asistimos al relanzamiento de un novedoso modo de construir y validar instrumentos de medida basado en los modelos de ecuaciones estructurales (SEM como se les conoce por su acrónimo en inglés). Se trata de técnicas estadísticas multivariantes que combinan la regresión múltiple y el análisis factorial al servicio de la validación de teorías, su cuantificación y la dirección de las relaciones de causalidad entre variables. Permite pues, contrastar la validez de una teoría a partir de los datos obtenidos mediante diferentes instrumentos, a los que se evalúa en sus características psicométricas. Dos son las principales ventajas que se señalan de los modelos SEM: la posibilidad de trabajar con variables latentes no medibles directamente y poder considerar los errores de medida como parte del modelo.

En estos modelos se trabaja en un doble sentido: evaluando el modelo de medida y comprobando el modelo estructural o predictivo. Este último establece las relaciones entre variables latentes y el resto de relaciones que no forman parte del modelo de medida. Su objetivo final es obtener los

pesos de las relaciones entre variables (la fuerza de sus relaciones causales) y la naturaleza de la relación (directa o inversa).

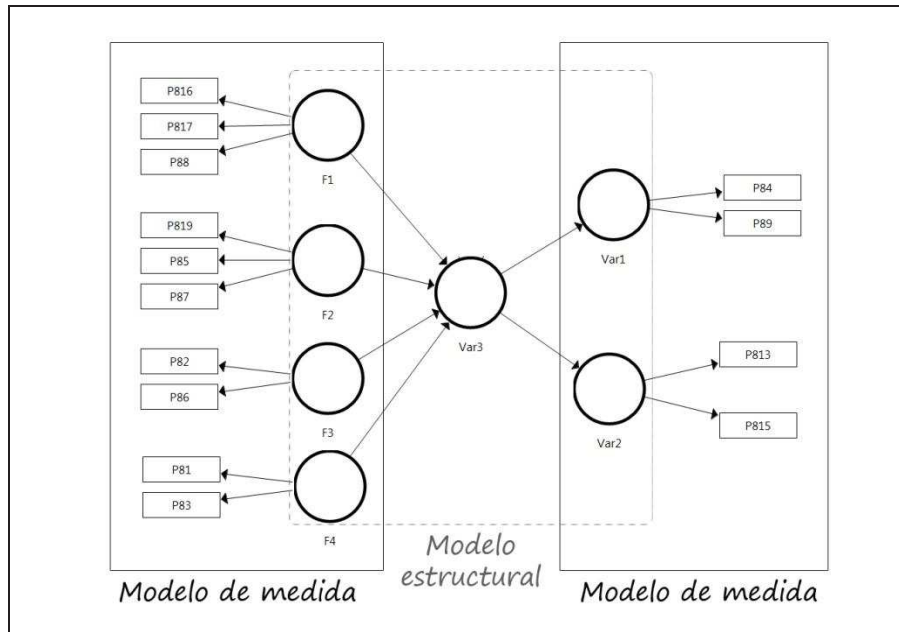


Figura 2.- Modelo estructural y modelo de medida de un análisis SEM

En el modelo de medida se especifican las relaciones de los indicadores y sus factores latentes asociados. Dado que, como hemos indicado, las variables latentes no pueden medirse directamente, debemos suponer que se manifiestan a través de sus indicadores. La bondad de la medición y, por tanto, su validez psicométrica, dependerá de la relación real que pueda existir entre los indicadores observables (las respuestas a nuestros ítems) y los constructos subyacentes. Si esta relación es débil o incluso inadecuada, la medición será imprecisa o simplemente incorrecta. Resumimos en la Tabla 3 algunos de los indicadores de ajuste más utilizados.

Tabla 3. Indicadores del ajuste de un modelo de medida (análisis SEM)

Indicador	Descripción	Valores admisibles
Nivel de significación de χ^2	Este estadístico permite contrastar la hipótesis nula de que el modelo es correcto.	El valor de p (χ^2) debería ser superior a .05
Razón χ^2 /gl	Permite contrastar la hipótesis nula de que el modelo es correcto en razón de los grados de libertad del modelo.	Debería ser inferior a 2.00
NNFI	Índice de ajuste no normado de Tucker y Lewis (1973)	Debería ser superiores a .95; mejor cuanto más próximo a 1.00
CFI	Índice de ajuste comparado de Bentler (1990)	Debería ser superiores a .95; mejor cuanto más próximo a 1.00
RMSEA	Error cuadrático medio de aproximación	Inferior a .08 (preferiblemente, inferior a .06); el modelo debería rechazarse si $RMSEA > .10$
SRMR	Valor estandarizado del índice RMSEA obtenido al dividir dicho valor por la desviación típica	Inferior a .08, mejor mientras más próximo a .00
Valores de t	Grado de significación estadística de las relaciones entre las variables intervinientes	Los valores absolutos deberían ser superiores a 1.96

en el modelo

Saturaciones	Valor de las cargas de las relaciones del modelo	Superiores a .30
Residuos	Errores del modelo predicho	Distribución normal, simétrica en torno a 0, pocos residuos superiores a 2.00
Fiabilidad compuesta	Índice de fiabilidad que permite tener en cuenta todos los constructos implicados en la escala, y no un análisis uno a uno como Cronbach	Las fiabilidades compuestas de las variables latentes deberían ser superiores a .60.
Varianza media extractada (VME)	Cantidad de varianza que un constructo obtiene de sus indicadores con relación a la cantidad de varianza debida al error	Las VME de las variables latentes deberían ser superiores a .50
Validez discriminante	La validez discriminante es grado en el que la varianza extractada de cada variable latente es superior al cuadrado de la correlación entre ellas; prueba que los constructos que no deberían tener ninguna relación de hecho, no la tienen.	Prueba de diferencias de χ^2 , los intervalos de confianza y la varianza extractada

Hablar de modelos de medida es referirse al Análisis Factorial Confirmatorio (AFC), procedimiento de análisis que se ha convertido en los últimos años en uno de los más usados en la validación de escalas de medida. Se trata de un procedimiento encuadrado dentro de los modelos SEM cuyo principal objetivo es analizar las relaciones entre un conjunto de indicadores observables y una o más variables latentes o factores.

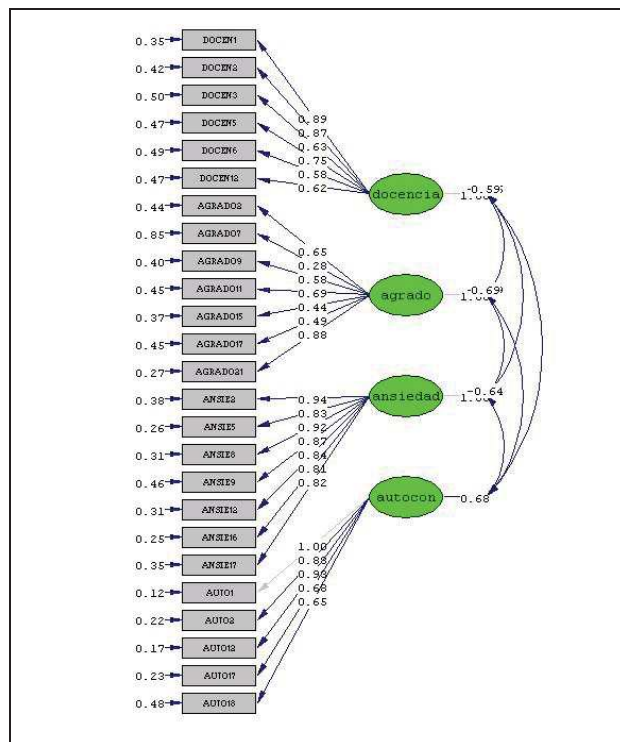


Figura 3.- Modelo de medida de un Análisis Factorial Confirmatorio (AFC)

Aunque se trata de técnicas complementarias al Análisis Factorial Exploratorio (AFE) es bien diferentes al AFC tanto por los objetivos que se persiguen como por los cálculos que se realizan. Como señalan Martínez, Hernández y Hernández (2006), la primera y más notoria diferencia se deduce de su propio nombre dado que en el AFE se trata de explorar la existencia de una no conocida estructura factorial de antemano, mientras que en el AFC se realiza una comprobación de una estructura conocida a priori mediante la formulación de un modelo de medida que se elabora previamente y se comprueba posteriormente. De ello se deriva, además, que en el AFE se desconoce inicialmente el número de factores, cosa que no sucede con el AFC que sabemos cuáles y cuántos de ellos intervienen en el modelo. Aspectos que pueden ser resumidos siguiendo a Hair, Anderson, Tatham y Black (1999) en que el AFE es una técnica multivariante especialmente indicada para el estudio de las relaciones de *interdependencia*, mientras que el AFC se enmarca dentro de los modelos SEM como técnicas multivariantes al servicio del estudio de las relaciones de *dependencia* entre diferentes variables.

Pese al valor que para la investigación tienen estos modelos de medida, el AFC es una técnica de análisis exigente pues son numerosas las condiciones necesarias para su realización (Tabla 4).

Tabla 4.- Condiciones necesarias para realizar una AFC (tomado de Arias, 2008)

<i>Condición</i>	<i>Observaciones</i>
1. Nivel de medida	Indicadores en nivel de intervalo o de razón (excepcionalmente, ordinal)
2. Valores por indicador	Los indicadores deberían tener un mínimo de 4 valores.
3. Normalidad y <i>outliers</i>	Distribución normal de los datos, control de outliers.
4. Homocedasticidad	Corrección mediante normalización o transformaciones.
5. Datos perdidos	Tratamiento adecuado de los datos perdidos.
6. Tipo de relaciones	Relaciones lineales y aditivas.
7. Multicolinealidad	Ausencia de multicolinealidad
8. Variables relevantes	Inclusión dentro del modelo de todas las variables relevantes.
9. Identificación del modelo	Modelo supraidentificado.
10. Número mínimo de observaciones	Al menos 150 observaciones, o 5 observaciones por cada parámetro a estimar.
11. Indicadores por variable latente	Preferible disponer de más de 2 (lo ideal es disponer de al menos 4 o 5).
12. Número de indicadores	El número máximo de indicadores no debería exceder de 20-30.
13. Varianzas relativas	Ausencia de matrices <i>ill-scaled</i>

Para solucionar al menos en parte estas exigencias han surgido nuevos modelos que podríamos denominar *no paramétricos* que mantienen los principios de los modelos SEM. Los modelos basados en covarianzas (CB SEM), como son los que se basan en el algoritmo del programa Lisrel de Jöreskog (1970), están sustentados en un enfoque muy paramétrico de máxima verosimilitud

(ML) que plantea muchas restricciones respecto al tamaño de la muestra y las propiedades que deben tener los datos del modelo. Sin embargo, Herman Wold, que fuera director de tesis del propio Jöreskog, consideró que era necesario un sistema menos exigente y más de acuerdo al tipo de datos con los que habitualmente se trabaja en ciencias sociales. Este planteamiento, en palabras de Tenenhaus y Vinzi (2005) contraponía un enfoque que denominó *soft modelling* frente al enfoque *hard modelling* de Jöreskog. Este nuevo planteamiento lo denominó *Partial Least Squares* (PLS). Estos modelos aúnan dos técnicas multivariantes cuales son el análisis de componentes principales y la regresión lineal múltiple (Figura 4).

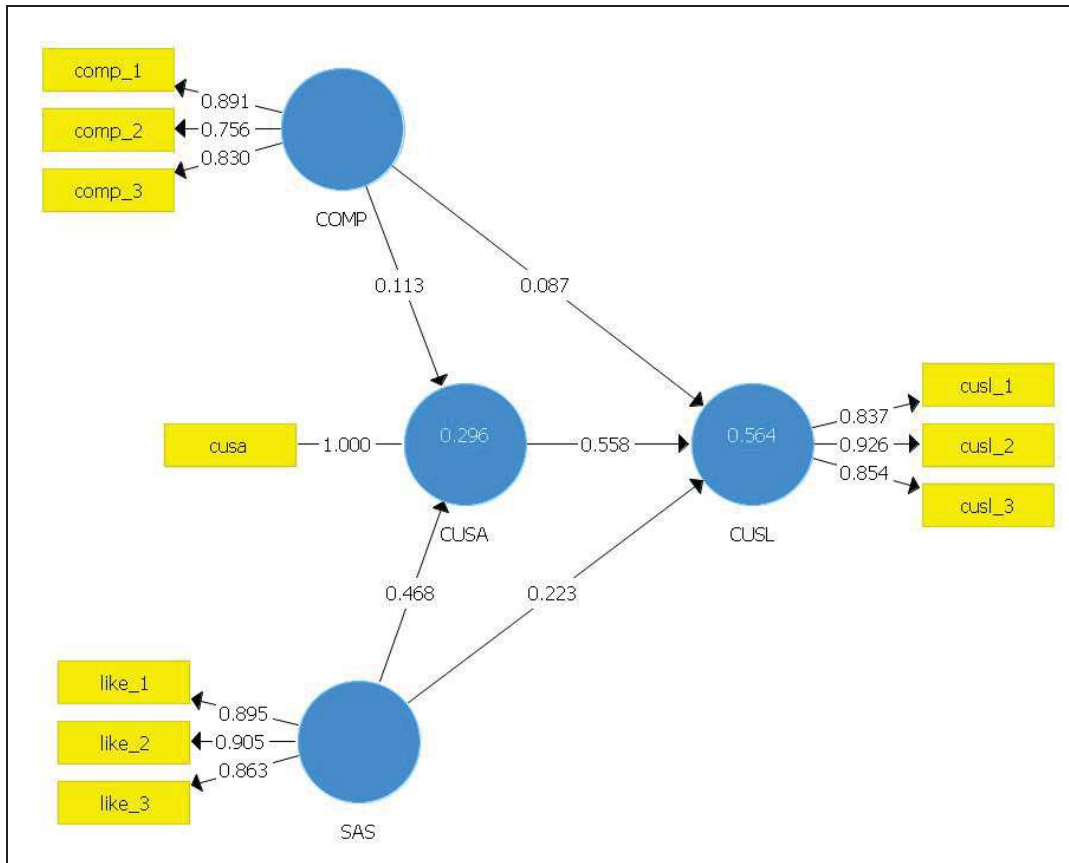


Figura 4.- Modelo de medida y modelo estructural en PLS

Entre sus ventajas cabe señalar la capacidad de identificación del modelo con tamaños de muestras pequeños, con las que, por lo general, se alcanzan niveles altos de potencia estadística, no se necesitan asunciones sobre las distribuciones de los datos, son poco sensibles a la presencia de valores perdidos, admiten variables latentes medidas con uno o varios indicadores, incorporan con facilidad modelos de medida tanto formativos como reflectivos y manejan sin problemas modelos complejos con gran cantidad de relaciones estructurales.

Como el resto de modelos SEM, en los modelos PLS es posible diferenciar la parte relativa a la medida del modelo estructural propiamente dicho. Los modelos de medida posibilitan evaluar la fiabilidad y la validez (Figura 4).

LA MEDIDA DEL DOMINIO AFECTIVO MATEMÁTICO

La medida del dominio afectivo en matemáticas no ha sido indiferente al devenir histórico de los modelos que hemos intentado resumir en el apartado anterior. Así, aunque gran parte de las escalas instrumento siguieron el modelo de las Teorías Clásicas de Medida, se está imponiendo con cada vez más fuerza los estudios que siguen las Teorías de Respuesta al Ítem y, sobre todo, las propuestas basadas en los modelos de ecuaciones estructurales.

En las líneas que siguen, realizamos un breve recorrido por los instrumentos más representativos, desde estas diferentes posturas, en relación a la medida de dos de los tres componentes del dominio afectivo matemático señalados por McLeod (1988, 1992) cuales son las actitudes hacia las matemáticas y la ansiedad hacia las matemáticas.

Se ha señalado sobradamente el importante papel que las *actitudes hacia las matemáticas* tienen en el proceso de enseñanza-aprendizaje y sobre el rendimiento matemático de los alumnos (Miñano y Castejón, 2011; Miranda, 2012; Sakiz, Pape, y Hoy, 2012). Asimismo, es de sobra conocida la influencia negativa que las buenas actitudes hacia las matemáticas tienen sobre la ansiedad (Akin y Kurbanoglu, 2011). Otros estudios han encontrado que los estudiantes con mejores actitudes hacia las matemáticas perciben las matemáticas como más útiles y muestran mayor motivación intrínseca por esta materia (Perry, 2011), poseen autoconceptos matemáticos más elevados (Hidalgo, Maroto, y Palacios, 2005), mayor confianza en su aprendizaje (McLeod, 1992) y muestran conductas de acercamiento a esta materia (Fennema y Sherman, 1976). La importancia del estudio de las actitudes explica el interés de su medida. Para comprender una creencia o un estado emocional, su posible cambio o la correcta predicción de una conducta, necesitamos de medidas válidas y fiables.

No obstante, esta medida plantea importantes problemas pues, por su propia naturaleza, esta medida es siempre indirecta. Las actitudes, lo mismo que las creencias sólo son medibles sobre la base de determinadas inferencias sobre sus componentes: sobre sus acciones explícitas, sus contenidos o sus estados emocionales. Las actitudes son indicadores de conducta, pero no la conducta en sí. En este sentido, se han encontrado discrepancias importantes entre las actitudes y la conducta, siendo el fenómeno que se conoce con el nombre de *deseabilidad social* un importante sesgo que puede explicar dichas discrepancias.

Para evitar o disminuir este sesgo, se han desarrollado técnicas de medida camufladas como alternativa a las escalas directas. En este tipo de medidas indirectas, se utilizan sistemas de enmascaramiento al medir sutilmente los pensamientos y los afectos que se relacionan con el objeto de estudio. Cabe señalar como ejemplos de estos métodos indirectos de medida los registros fisiológicos (electroencefalogramas o electromiografía facial), las pruebas proyectivas, los métodos de observación conductual, las tareas de evaluación automática o los Test de Asociación Implícita.

Pese al valor de estos procedimientos, son más habituales las medidas directas de las actitudes a partir de procedimientos como el diferencial semántico, las escalas de intervalos, o las escalas Likert, procedimientos que suelen englobarse en la categoría de sistemas de autoinforme o simplemente escalas de actitud (Tabla 5).

Tabla 5.- Técnicas de autoinforme en la medida de las actitudes

Técnica	Autor/es	Descripción
Escala de Diferencial Semántico	Osgood, Suci y Tannenbaum (1957)	Los encuestados evalúan un objeto según una escala en la que se evalúa mediante dos conceptos contrapuestos y bipolares separados por intervalos
Escala de Intervalos	Trurstone (1928)	Se mide la actitud mediante un continuo de ítems escalonados que, previamente, han preparado los

		investigadores
Escala Likert	Likert (1932)	Conjunto de proposiciones-afirmaciones sobre las que hay que posicionarse generalmente según el grado de acuerdo medido en una escala de diferentes puntos
Escalograma de Guttman	Guttman (1947)	Se trata de un procedimiento que se centra en la determinación de las propiedades de un conjunto de ítems.

Las escalas de intervalos o diferenciales de Thurstone son consideradas como el primer método sistemático para medir actitudes. Este tipo de escalas están construidas en base a un conjunto de preguntas en forma de afirmaciones que varían según su mayor o menor grado de acuerdo con respecto o al objeto que se pretende medir. Este grado de favorabilidad de las preguntas se determina previamente mediante las opiniones de un grupo de jueces que evalúan esa característica. La construcción de este tipo de escalas es muy laborioso por lo que se suele acudir a otro tipo de escalas como puedan ser las sumativa tipo Likert.

Estas escalas Likert son las más conocidas y utilizadas, entre otros aspectos, por la sencillez de su construcción. Se fundamenta en un sistema utilizado en otro tipo de instrumentos de evaluación cual es: situar a los individuos a lo largo de un continuo a partir de la suma de las respuestas a diferentes preguntas que deben poseer el mayor grado de homogeneidad; razón por la cual, también se las denomina escalas sumativas. Es importante resaltar que se debe conseguir el mayor grado de parecido entre los ítems, dado que debemos suponer que todos ellos miden lo mismo (supuesto de unidimensionalidad de la escala no siempre comprobado). Para evitar tendencias de respuestas no deseadas, se deben incluir preguntas enunciadas en términos favorables y preguntas en términos desfavorables.

Por lo general, todas ellas tienden a centrarse en la medida de la valencia y la intensidad de las creencias y conocimientos así como de los sentimientos del encuestado y la tendencia actuar de manera determinada de manera conjunta o por separado.

Las propuestas para medir las actitudes hacia las matemáticas surgen tempranamente siendo los trabajos de Aiken (Aiken, 1972, 1974, 1979; Aiken y Dreger, 1961) y Dutton y Blum (1968) pioneros en el tiempo. El trabajo de Aiken y Dreger (1961) propone un cuestionario compuesto por 20 ítems con dos subescalas: *Agrado y Miedo a las matemáticas*, aunque algunos autores la han considerado como una escala unidimensional (Auzmendi, 1992). En una versión posterior, Aiken (1972) introduce el factor *Disfrute de las Matemáticas*, sobre la que, años más tarde, añadirá dos subescalas: *Valor de las matemáticas* y *Disfrute de las matemáticas*. En una versión posterior, Aiken (1979) aumentará el número de factores hasta un total de cuatro: *Gusto por las matemáticas*, *Motivación matemática*, *Valor-Utilidad de las matemáticas* y *Miedo a las matemáticas*.

La escala de Fennema y Sherman (1976) es para muchos la más popular de las medidas de las actitudes hacia las matemáticas (Tapia y Marsh, 2004). Esta escala ha sido objeto de amplios estudios de replicación, traducida a diferentes lenguas y modificada para ser aplicada a diferentes situaciones. La aportación de Tapia y Marsh (2004) denominada ATMI (Attitude Toward Mathematics Inventory) es, sin duda, uno de los instrumentos más utilizados en la medida de las actitudes hacia las matemáticas. La versión final está formada por 49 ítems que miden seis factores: *Confianza-autoconcepto*, *Ansiedad*, *Utilidad-valor de las matemáticas*, *Gusto por las matemáticas*, *Motivación* y *Expectativa de los padres y profesores*.

En todas estas propuestas, el modelo psicométrico seguido es el de la Teoría Clásica de medida presentando como indicadores de fiabilidad exclusivamente índices de homogeneidad de los ítems

(alfa de Cronbach) y análisis factoriales exploratorios (AFE) como criterio único de validez como medio de conocer y describir la estructura factorial sin que se realicen análisis factoriales confirmatorios posteriores (AFC).

No conocemos adaptaciones propiamente dichas al castellano de las escalas de Aiken (1974), Fennema y Sherman (1976) y Tapia y Marsh (2004). En la mayor parte de las investigaciones en nuestra lengua se han limitado a su uso sin que exista estudio previo de las propiedades psicométricas de estas escalas. Señalamos entre otros trabajos los realizados por por Cazorla, Silva, Vendramini, y Brito (1999) de la escala de Aiken (1974) sobre la base de una anterior al portugués de Brito (1998) orientada al estudio de las actitudes hacia la estadística, la más moderna de Estrada y Díez-Palomar (2011) o la de González-Pienda et al. (2012).

En la medida de las actitudes hacia las matemáticas en lengua castellana puede considerarse pionero el trabajo de Gairín (1990). Aunque es Auzmendi (1992) quien elabora la que es, sin duda, la escala de actitudes hacia las matemáticas más citada de las realizadas en lengua castellana (Palacios, Arias y Arias, 2014). Más cercanos en el tiempo son los trabajos de de Muñoz y Mato (2008) y de Alemany y Lara (2010). La escala de Muñoz y Mato (2008) consta de 19 ítems que, tras el análisis factorial correspondiente, presenta dos únicos factores: Actitud del profesor percibida por el alumno y Agrado-utilidad de las matemáticas. La aportación de Alemany y Lara (2010), trabajando con una muestra de validación de alumnos de origen étnico berebere, presenta un alfa de Cronbach de .92 en una muestra de 236 estudiantes de 2.º y 3.º de Educación Secundaria.

Como señalan Palacios, Arias y Arias (2014), las diferentes escalas de actitudes hacia las matemáticas tanto en inglés como en castellano que brevemente hemos analizado presentan, en general, índices de fiabilidad adecuados, si no se tienen en cuenta las limitaciones del coeficiente alfa de Cronbach utilizado para evaluar dicha fiabilidad.

Existen pocas propuestas de medida de las actitudes hacia las matemáticas basadas en los modelos TRI, SEM o PLS. Entre ellas, queremos señalar la aportación de Palacios, Arias y Arias (2014) pues, además de la cercanía en el tiempo, se trata de una propuesta de una escala en castellano, con una muestra elevada de sujetos (4741 participantes), que resume el modo de trabajo de estos nuevos modelos de análisis de medida. Para la elaboración de su *Escala de Actitudes hacia las Matemáticas* (EAM) parten los autores de trabajos anteriores como los señalados en este apartado que presentan cinco factores de manera generalizada en el constructo actitudes hacia las matemáticas: *agrado-gusto por las matemáticas, ansiedad hacia las matemáticas, percepción de dificultad, utilidad percibida y autoconcepto matemático*. Tras un primer AFE, los autores concluyen que la estructura subyacente de la prueba está formada por cuatro factores: *Percepción de la incompetencia matemática, Gusto por las matemáticas, (Percepción de utilidad) y Autoconcepto matemático*. Como paso lógico dentro de estos nuevos modelos, se presentan los resultados de diferentes AFC los que presentan como mejor modelo el de cuatro factores correlacionados (Figura 5).

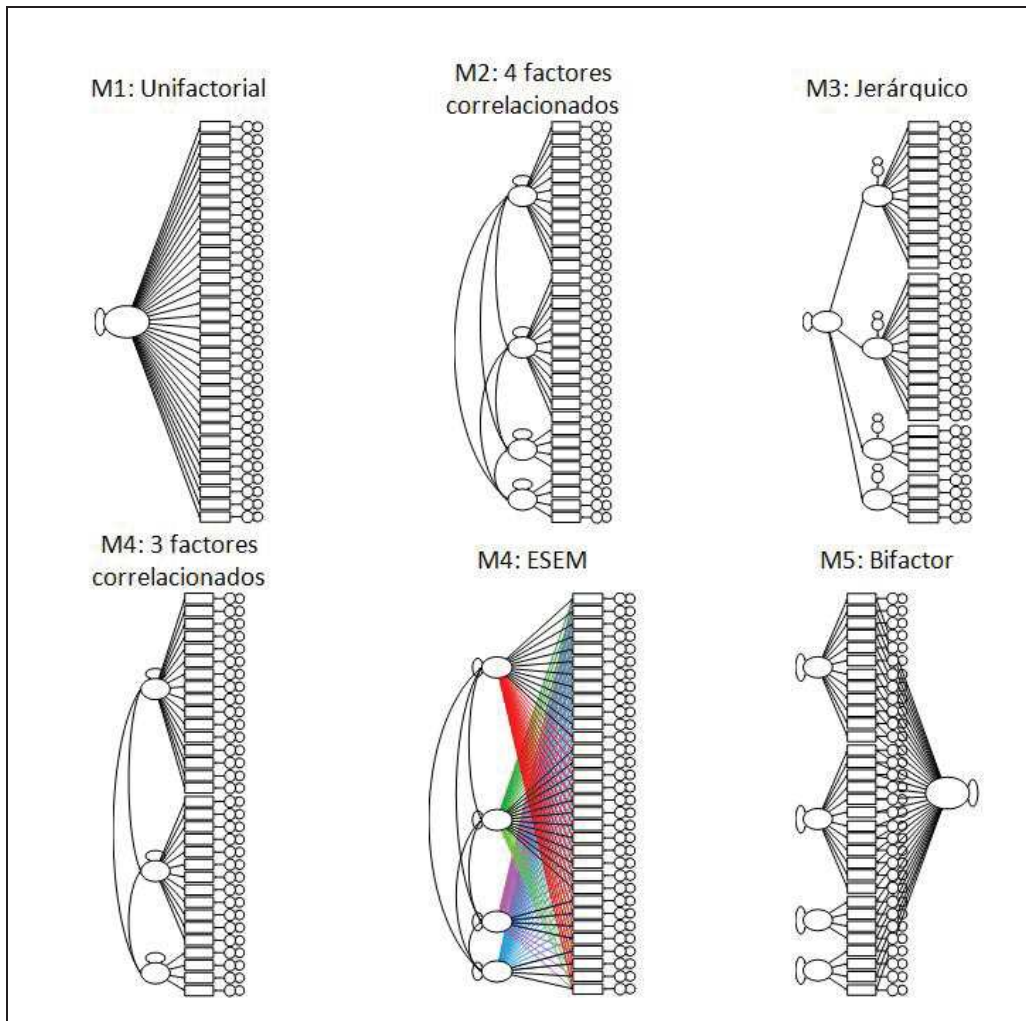


Fig. 5.- Contrastación mediante AFC de los modelos de la escala de actitudes de Palacios, Arias y Arias (2014)

Dentro de este mismo modelo de TRI y como alternativa al cálculo de la fiabilidad mediante el alfa de Cronbach, los autores calculan la fiabilidad compuesta (*composite reliability*) a partir de las saturaciones y los errores de medida, el coeficiente alfa ordinal, el coeficiente theta ordinal y el coeficiente Omega de McDonald y el *glb* (*greatest lower bound*). Se complementan estos valores con los cálculos de la *validez de contenido*, la *validez de constructo* y la *validez convergente*. Como concluyen los propios autores, la escala presenta evidencias tanto de validez como de fiabilidad. Estas evidencias, junto con la garantía de haber sido obtenidas con una muestra de gran tamaño (con las consiguientes garantías de potencia estadística y disminución del error de medida) y los modelos SEM señalados permiten concluir a los autores que se trata de un instrumento de medida de las actitudes hacia las matemáticas sólido y robusto, y de una gran utilidad potencial para su uso en niveles educativos no universitarios.

La ansiedad hacia las matemáticas es otro de los componentes básicos de las emociones y por ende, del dominio afectivo matemático. En una de las definiciones clásicas, Richardson y Suinn (1972) describen la ansiedad hacia las matemáticas como “*a feeling of tension and anxiety that interfere with the manipulation of mathematics problems in varied situations in ordinary as well as academic life*”. Como sucediera con las actitudes, por su trascendencia, pronto aparecieron escalas para su evaluación. Hay que considerar el trabajo de Richardson y Suinn (1972) como el primer instrumento diseñado específicamente para medir la ansiedad matemática. Esta escala, denominada *Mathematics Anxiety Rating Scale* (MARS), estaba formada por 98 preguntas con cinco alternativas

de respuesta. En sus orígenes, fue construido como medida unidimensional de la ansiedad matemática. La escala MARS ha sido adaptada a diferentes poblaciones y edades. Suinn y Edwards (1982) validaron la *Mathematics Rating Scale for Adolescents* (MARS-A) para una población de 16 a 25 años, que posteriormente la adaptarían para estudiantes de primaria conocida como la escala MARS-E (*Mathematics Anxiety Rating Scale for Elementary School*) (Suinn, Taylor y Edwards, 1988). Las escalas basadas en MARS han sido adaptadas a diferentes idiomas. Nuñez-Peña, Suarez-Pellicioni, Guilera y Mercadé-Carranza (2013) realizaron una adaptación al castellano; Mahmood y Khatoon (2011) al idioma indio; Ko y Yi (2011) al coreano; al turco (Akin, Kurbanoglu y Takunyaci, 2011; Baloglu, 2010); al iraní (Vahedi y Farrokhi, 2011); al idioma italiano (Primi, Busdraghi, Tomasetto, Morsanyi y Chiesi, 2014)

No obstante, como han sugerido Baloglu y Zelhart (2007), pronto quedaron patentes dos problemas de la MARS: a) la duración y longitud exagerada de la prueba b) basarse en un supuesto no suficientemente comprobado de unidimensional de la ansiedad matemática. Entre los primeros intentos de realizar una versión reducida hay que mencionar la propuesta de Plake y Parker (1982) de la escala MARS-R (*Mathematics Rating Scale Revised*), compuesta por 24 ítems. En esta misma línea de versiones reducidas, Suinn y Winston (2003) elaboraron la que se conoce con el nombre de MARS-SV (*Mathematics Anxiety Rating Scale Short Version*), compuesta por 30 preguntas y una estructura unidimensional parecida a la obtenida con la versión original de 98 ítems. Más cercana en el tiempo, cabe señalar la escala Cipora, Szczygiel, Willmes y Nuerk (2015).

La versión más reducida de la escala es sin duda la de Nuñez-Peña, Guilera y Suarez-Pellicioni (2013) quienes propusieron un instrumento compuesto por un solo ítem (“On a scale from 1 to 10, how math anxious are you?”) denominada por los autores SIMA (*Single Item Math Anxiety*). Aunque cabe señalar al trabajo de Alexander y Martray (1989) como una de las adaptaciones más utilizadas entre las versiones reducidas de MARS.

Además de la citada MARS, cabe señalar la escala MAS (*Mathematics Anxiety Scale*) de Fennema y Sherman (1976) en su adaptación de Betz (1978) como otro de los referentes de la medida de la ansiedad matemática. La MAS está pensada para la evaluación de la ansiedad matemática a partir de un conjunto de 10 ítems en una escala Likert de cinco puntos. Bai, Wang, Pan y Frey (2009) aumentaron a 14 los ítems de la escala anterior y la aplicaron a una muestra de estudiantes universitarios. Esta versión presentó índices de fiabilidad mejores que la versión original de la escala. Las escalas MAS han sido traducidas también a diferentes idiomas. Concretamente, Hunt, Clark-Carter y Sheffield (2011) adaptaron la escala original a una población del Reino Unido denominada MAS-UK (*Mathematics Anxiety Scale for British Undergraduates*). Lim y Chapman (2013) realizaron una nueva versión compuesta por 9 ítems con una muestra de estudiantes pre-universitarios de la ciudad de Singapur, la FSMAS-R; Prieto y Delgado (2007) la adaptaron para evaluar estudiantes españoles.

Se han desarrollado igualmente escalas para niños de las que cabe señalar el trabajo de Chiu y Henry (1990) con un instrumento de 22 ítems con adecuada validez y fiabilidad, y una estructura de cuatro factores: *Math Evaluation, Learning, Problem Solving y Teacher Anxiety*. Dentro de este mismo tipo de pruebas para niños, señalamos la adaptación de la escala MARS-E de Suinn, Taylor y Edwards (1988) y Ramirez, Chang, Maloney, Levine y Beilock (2016) denominada *Revised Child Math Anxiety Questionnaire* (CMAQ-R). Además, cabe señalar la CAMS (*Children's Anxiety in Math Scale*; Jameson, 2013). Está formada por 16 ítems con una estructura de tres factores (*General Math Anxiety, Math Performance Anxiety y Math Error Anxiety*) y presenta índices de fiabilidad adecuados. También, cabe señalar por su importancia el trabajo de Wigfield y Meece (1988), quienes elaboraron la escala MAQ (*Math Anxiety Questionnaire*), especialmente diseñada para los estudiantes de primaria y secundaria. Se trata de una adaptación compuesta por 11 ítems con una estructura de dos factores (*Negative Affective Reactions y Worry*), una adecuada fiabilidad y unas correlaciones en las direcciones esperadas con el rendimiento matemático.

Como sucediera con las actitudes, la mayor parte de las escalas de medida de la ansiedad matemática fueron validadas según el modelo clásico siendo escasas las que se ajustaron a los modelos más modernos de TRI o modelos estructurales. Cabe considerar como mejor ejemplo de estas últimas los trabajos de Palacios et al. (2013), Prieto y Delgado (2007) y Nuñez-Peña et al. (2013).

En el trabajo de Palacios et al (2013) se contrasta un modelo de ecuaciones estructurales en el que, una parte importante del trabajo, se dedica a la comprobación de la fiabilidad y la validez de una escala de ansiedad hacia las matemáticas (modelo de medida). La escala final compuesta por 16 ítems, que se presenta como una adaptación al contexto español de la escala MARS mencionada de Richardson y Suinn (1972), presenta valores de ajuste adecuados con una solución unidimensional tras el AFC (RMSA=.045; CFI=.97; AGFI=.95; NFI=.96). En un estudio anterior, Prieto y Delgado (2007) realiza la validación al contexto español de la escala MAS (Math Anxiety Scale) mediante la teoría del TRI y concretamente del modelo Rasch. La validez convergente obtenida mediante la correlación con la escala de actitudes de Fennema y Sherman (1976) es elevada, presentando la escala propiedades métricas óptimas, resultados que, a grandes rasgos, también obtienen Nuñez-Peña et al. (2013). En este caso, se analiza la escala sMARS de Alexander y Martray (1989) comprobando, mediante un AFC la estructura de tres factores de la escala original (*math test, numerical task y math course anxiety*). Presenta, además, una excelente consistencia interna de estas tres subescala y de la escala total, así como evidencias de una adecuada validez convergente y validez discriminante; lo que lleva a los autores a concluir que la adaptación al castellano de la escala sMARS posee propiedades psicométricas adecuada tanto por su consistencia interna como por su validez.

CONCLUSIONES

Los procedimientos de recogida de información y de medida son aspectos de vital importancia en la investigación en general y en la investigación matemática en particular. Aunque los métodos no deben determinar los objetivos de las investigaciones, los afecta de manera importante. Entre otras razones, porque si los instrumentos de recogida de datos son inadecuados, también lo serán sus conclusiones. Por ello, es necesario asegurar que estos instrumentos tengan propiedades psicométricas que garanticen la exactitud de la medida. Más si cabe pues, como hemos indicado, en educación matemática, como en la mayor parte de las ciencias sociales, medimos conceptos y rara vez cosas tangibles (constructo).

En este sentido, el conocimiento de las teorías estadísticas sobre la medida es de gran valor para la investigación matemática pues nos va a asegurar las principales cualidades psicométricas de nuestros instrumentos; nos va a permitir determinar su grado de fiabilidad y validez, aspectos ambos necesarios para usar los instrumentos con garantía, de forma rigurosa y científica.

La Teoría Clásica de Medida surge en gran medida como soporte estadístico de los test que tanto éxito comenzaron a tener a comienzos del siglo pasado. Su fundamento estadístico es simple: la puntuación verdadera de una escala es la suma de la puntuación obtenida por el sujeto (puntuación empírica) más un error. Este error será difícil de cuantificarlo pero posible de estimar ya que tiende a cero al pasar infinitas veces la misma prueba, al mismo sujeto. La fiabilidad del instrumento será entonces la capacidad que posea de minimizar la cuantía de este error. Un instrumento medirá con precisión si comete pocos errores en su medida.

Pero no basta saber que una medida es consistente si no estamos seguros de medir lo que pretendemos medir; planteamiento que nos acerca al concepto de validez de una medida. Pese a que su cálculo ocupó un espacio menos relevante en la teoría clásica de medida que la fiabilidad, se

propusieron procedimientos que aseguraran estas características de todo buen instrumento entre las que destacamos la validez de contenido, la validez de criterio o la validez constructo.

Además, un buen instrumento de medida debe conllevar un grupo normativo de referencia que nos permita establecer el significado de las puntuaciones empíricas. Este grupo normativo actúa como el referente que nos va a permitir situar a cada sujeto, en cuanto a los resultados de la prueba, en un lugar determinado (un percentil, por ejemplo) en comparación con este grupo normativo. Como cabe suponer, un mismo sujeto puede ocupar una posición diferente en función de diferentes grupos normativos.

Esta Teoría Clásica de Medida realiza un conjunto de presunciones de dudosa validez sino directamente erróneas. Es aventurado afirmar, por ejemplo, que la medida de la actitud hacia las matemáticas dependa del cuestionario que se utilice; asegurar que el resultado en el cuestionario está determinado por los sujetos a los que hemos pasado la escala en el grupo normativo. La TRI surge como un intento de solución de estos presupuestos difíciles de demostrar, para ello, intentarán realizar mediciones que no dependan del instrumento (invariante con respecto a la escala), ni de los grupos normativos (invariante respecto a las personas). No obstante, las estrictas condiciones que impone este modelo ha determinado la aparición de nuevas propuestas menos exigentes como los modelos basados en las ecuaciones estructurales (modelos SEM). En estos nuevos planteamientos, es posible realizar en un único modelo las cualidades de la medida (su fiabilidad y su validez) y el grado de ajuste de los datos a la teoría que lo sustenta. El concepto de *variable latente* juega entonces un papel prioritario y se pone de manifiesto en la realización de análisis factoriales confirmatorios (AFC) como alternativa a los análisis factoriales exclusivamente exploratorios (AFE).

La medida del dominio afectivo matemático no ha sido ajena al devenir histórico de estos modelos. Así, aunque gran parte de las escalas de medida del dominio afectivo matemático han seguido el modelo clásico, se impone actualmente con fuerza aportaciones novedosas basadas en los modelos de la TRI y de los modelos SEM. En este sentido, cabe recordar las aportaciones de Palacios, Arias y Arias (2014) en el campo de la medida de las actitudes hacia las matemáticas y los trabajos de Palacios et al. (2013), Prieto y Delgado (2007) y Nuñez-Peña et al. (2013) en la medida de la ansiedad hacia las matemáticas.

Referencias bibliográficas

- Aiken, L. R. (1972). Research on attitudes toward mathematics. *The Arithmetic Teacher*, 19(3), 229-234.
- Aiken, L. R. (1974). Two scales of attitude toward mathematics. *Journal for Research in Mathematics Education*, 5, 67-71.
- Aiken, L. R. (1979). Attitudes toward mathematics and science in Iranian middle schools. *School Science and Mathematics*, 79, 229-234
- Aiken, L. R., y Dreger, R. M. (1961). The effect of attitude on performance in mathematics. *Journal of Educational Psychology*, 52, 19-24.
- Akin, A., Kurbanoglu, N.I. y Takunyaci, M. (2011). Revised mathematics anxiety rating scale: a confirmatory factor analysis. *Journal of Science and Mathematics Education*. 5(1), 163-180.
- Akin, A., y Kurbanoglu, N. (2011). The relationships between math anxiety, math attitude and self-efficacy: A structural equation model. *Studia Psychologica*, 53(3), 263-273.

- Aleman, I., y Lara, A. I. (2010). Las actitudes hacia las matemáticas en el alumnado de la ESO: un instrumento para su medición. *Publicaciones*, 40, 49-71.
- Alexander, L., y Martray, C. (1989). The development of an abbreviated version of the Mathematics Anxiety Rating Scale. *Measurement and Evaluation in Counseling and Development*, 22(3), 143-159
- Arias, B. (2008). Desarrollo de un ejemplo de Análisis Factorial Confirmatorio con Lisrel, AMOS y SAS. En Verdugo, M. A., Crespo, M., Badía, M. y Arias (Coordinadores) (2008). *Metodología en la investigación sobre discapacidad. Introducción al uso de las ecuaciones estructurales*. Publicaciones del INICO: Salamanca
- Attorresi, H.F., Lozzia, G.S., Abal, F. Galibert, M. S. y Aguerri, M. E. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18, 179-188
- Auzmendi, E. (1992). *Las actitudes hacia la matemática-estadística en las enseñanzas medias y universitarias. Características y medición*. Bilbao: Mensajero.
- Bai, H., Wang, L., Pan, W., y Frey, M. (2009). Measuring mathematics anxiety: Psychometric analysis of a bidimensional affective scale. *Journal of Instructional Psychology*, 36, 185-193.
- Baloğlu, M. (2010). An investigation of the validity and reliability of the adapted mathematics anxiety ratings scale-short version (MARS-SV) among Turkish students. *European Journal Psychology Education*, 25, 507-518
- Baloglu, M. y Zelhart, P. (2007). Psychometric properties of the revised Mathematics anxiety rating scale. *The Psychological Record*, 57, 593-611
- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 238-246, 1990.
- Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Counselling Psychology*, 25, 441-448.
- Brito, M. R. F. (1998). Adaptação e validação de uma escala de atitudes em relação à matemática. *Zetetiké*, 6(9), 109-162.
- Cazorla, I. M., Silva, C. B., Vendramini, C., y Brito, M. R. F. (1999). Adaptação e validação de uma escala de atitudes em relação à estatística. *Actas de la Conferência Internacional Experiências e Perspectivas do Ensino da Estatística*. PRESTA, Florianópolis: Florianópolis.
- Cipora, K., Szczygiel, M., Willmes, K. y Nuerk, H.C. (2015). Math Anxiety Assessment with the Abbreviated Math Anxiety Scale: Applicability and Usefulness: Insights from the Polish Adaptation. *Frontiers in Psychology*, 6, 1-16.
- Chiu, L. H., y Henry, L. L. (1990). Development and validation of the Mathematics Anxiety Scale for Children. *Measurement and Evaluation in Counseling and Development*, 23, 121-127
- Dutton, W. H., y Blum, M. P. (1968). The measurement of attitudes toward arithmetic with a Likert-type test. *Elementary School Journal*, 68, 259-264.
- Estrada, A., y Díez-Palomar, J. (2011). Las actitudes hacia las Matemáticas. Análisis descriptivo de un estudio de caso exploratorio centrado en la Educación Matemática de familiares. *Revista de Investigación en Educación*, 9(2), 116-132.
- Fennema, E., y Sherman, J. A. (1976). Fennema-Sherman Mathematics Attitudes Scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education*, 7(5), 324-326

- Gairín, J. (1990). *Las actitudes en educación. Un estudio sobre la educación matemática*. Barcelona: Boixareu Universitaria.
- González-Pienda, J. A., Fernández-Cueli, M., García, T., Suárez, N., Fernández, E., Tuero-Herrero, E., y Helena da Silva, E. (2012). Diferencias de género en actitudes hacia las matemáticas en la enseñanza obligatoria. *Revista Iberoamericana de Psicología y Salud*, 3(1), 55-73.
- Guttman, L. (1947). The Cornell technique for scale and intensity analysis. *Educational Psychological Measurement*, 7, 247-279.
- Hair, J.F.; Anderson, R.E.; Tatham R.L. y Black, W.C. (1999): *Análisis multivariante*. 5ª ed. Madrid: Prentice Hall.
- Hambleton, R. K., Swaminathan, H., y Rogers, J. (1991). *Fundamentals of item response theory*. Beverly Hills, CA: Sage
- Hidalgo, S., Maroto, A., y Palacios, A. (2005). El perfil emocional matemático como predictor de rechazo escolar: relación con las destrezas y los conocimientos desde una perspectiva evolutiva. *Revista de Educación Matemática*, 17(2), 89-116.
- Hunt, T.U., Clark-Carter, D. y Sheffield, D. (2011). The development and part validation of a U.K. Scale for Mathematics Anxiety. *Journal of Psychoeducational Assessment*. 29(5), 455- 466.
- Jameson, M. M. (2013). The Development and Validation of the Children's Anxiety in Math Scale. *Journal of Psychoeducational Assessment*, 31(4), 391-395.
- Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, , 57, 239-251.
- Ko, H.K. y Yi, H.S. (2011). Development and validation of a mathematics anxiety scale for student. *Asia Pacific Education Review*. 12(4), 509-521
- Lawshe C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*. 28(4), 563-575.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55.
- Lim, S.Y. y Chapman, E. (2013). An investigation of the Fennema-Sherman mathematics anxiety subscale. *Measurement and Evaluation in Counseling and Development*. 46(1) 26 -37.
- Lord, F. M., y Novick, M. R. (1968). *Statistical theories of mental test scores*. New York: Addison-Wesley.
- Mahmood, S. y Khatoon, A. (2011). Development and validation of the mathematics anxiety scale for secondary and senior secondary school students. *British Journal of Arts and Social Sciences*. 2(2), 169-179
- Martínez, R., Hernández, MV y Hernández, MJ. (2006). *Psicometría*. Madrid: Alianza.
- Martínez-Arias, M. R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid: Síntesis.
- McLeod, D. B. (1988). Affective issues in mathematical problem solving: Some theoretical considerations. *Journal for Research in Mathematics Education*, 19, 134-141
- McLeod, D. B. (1992). Research on affect in mathematics education: a reconceptualization. En D. Grows(Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575-596). New York: McMillan Publishing Company.

- Miñano, P., y Castejón, J. L. (2011). Variables cognitivas y motivacionales en el rendimiento académico en Lengua y Matemáticas: un modelo estructural. *Revista de Psicodidáctica*, 16(2), 203-230.
- Miranda, A. (2012). Funcionamiento ejecutivo y motivación en tareas de cálculo y solución de problemas de niños con TDAH. *Revista de Psicodidáctica*, 17(1), 51-72.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.
- Muñiz, J. (2010). La teoría de los test: Teoría clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, 31 (1), 57-66
- Muñoz, J. M., y Mato, M. D. (2008). Análisis de las actitudes respecto a las matemáticas en alumnos de la ESO. *Revista de Investigación Educativa*, 26(1), 209-226.
- Nuñez-Peña, M.L., Guilera, G. y Suarez-Pellicioni, M. (2013) The Single-Item Math Anxiety Scale: An Alternative Way of Measuring Mathematical Anxiety. *Journal of Psychoeducational Assessment* , 20(10), 1-12
- Nuñez-Peña, M.L., Suarez-Pellicioni, M. Guilera, G y Mercadé-Carranza, Cl. (2013) A Spanish version of the short Mathematics Anxiety Rating Scale (sMARS). *Learning and individual differences*. 24, 204-210
- Osgood, C.E., Suci, G.J., y Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.
- Palacios, A.; Arias, V., y Arias, B. (2014). Las actitudes hacia las matemáticas: construcción y validación de un instrumento para su medida. *Revista de Psicodidáctica*. 19(1). 67-91
- Perry, C. A. (2011). Motivation and attitude of preservice elementary teachers toward mathematics. Morehead State University. *School Science and Mathematics*, 111(1), 2-10.
- Plake, B. S., y Parker, C. S. (1982). The development and validation of a revised version of the Mathematics Anxiety Rating Scale. *Educational and Psychological Measurement*, 42, 551-557.
- Prieto, G., y Delgado, A. R. (2007). Measuring math anxiety (in Spanish) with the Rasch Rating Scale Model. *Journal of Applied Measurement*. 8, 149–160
- Primi, C., Busdraghi, C., Tomasetto, C., Morsanyi, K. y Chiesi, F. (2014). Measuring math anxiety in Italian college and high school students: validity, reliability and gender invariance of the Abbreviated Math Anxiety Scale (AMAS). *Learning and Individual Differences*. 34, 51–56.
- Ramirez, G., Chang, H., Maloney, E.A., Levine, S. C. y Beilock, S. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology*, 141, 83–100
- Richardson, F.C. y Suinn, M. (1972). The Mathematics Anxiety Rating Scale: Psychometric Data. *Journal of Counseling Psychology*. 18(6), 551-554
- Sakiz, G., Pape, S. J., y Hoy, A. W. (2012). Does perceived teacher affective support matter for middle school students in mathematics class rooms? *Journal of School Psychology*, 50, 235-255.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology*, 18, 161-169.

- Spearman, C. (1913). Correlations of sums and differences. *British Journal of Psychology*, 5, 417-426.
- Suinn, R. M., y Edwards, R. (1982). The measurement of mathematics anxiety: The mathematics anxiety rating scale for adolescents— MARS-A. *Journal of Clinical Psychology*, 38, 576-580.
- Suinn, R., Taylor, S., y Edwards, R. (1988). Suinn Mathematics Anxiety Rating Scale for elementary school students (MARS-E). Psychometric and normative data. *Educational and Psychological Measurement*, 48, 979-986
- Suinn, R. M., y Winston, E. H. (2003). The mathematics anxiety rating scales, a brief version: Psychometric data. *Psychological Reports*, 92(1), 167-173.
- Tucker, L.R. y Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 1-10, 1973
- Tapia, M., y Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, 8(2).
- Tenenhaus, M., y Vinzi, E.V. (2005). PLS regression, PLS path modeling and generalized procrustean analysis: a combined approach for PLS regression, PLS path modeling and generalized multiblock analysis. *Journal of Chemometrics*, 19, 145-153.
- Thurstone, L. L. (1928). Attitudes can be measured. *The American Journal of Sociology*, 26,249-269
- Vahedi, S., y Farrokhi, F. (2011). A confirmatory factor analysis of the structure of abbreviated math anxiety scale. *Iranian journal of psychiatry*, 6(2), 47.
- Wigfield, A. y Meece, J.L. (1988). Math Anxiety in Elementary and Secondary School Students. *Journal of Educational Psychology*, 80(2), 210-216