



## ANÁLISIS DE LAS PROPIEDADES PSICOMÉTRICAS DE UNA ESCALA DE ACTITUD: COMPARACIÓN DE LAS TÉCNICAS LIKERT Y THURSTONE.

Bernardo Alvarez Rodríguez  
Marcelino Cuesta  
Raquel Díaz Ramila  
Juan Manuel Jiménez García  
Dolores Paz Caballero  
Facultad de Psicología  
Universidad de Oviedo  
Plaza Feijóo s/n  
33003 Oviedo - España  
e-mail:pazcabal@correo.uniovi.es

### ABSTRACT.

The aim of this work was to compare “Thurstone” and “Likert” techniques for scale construction. To achieve this goal, two scales were elaborated from both the same item set and sample of subjects and then scored and analysed according to either Likert’ or Thurstone’ procedure. Thus, thirty items were chosen for each scale, eighteen of them were shared. Reliability coefficients ( $\alpha$ ) were higher for Likert-scale, even when responses were dichotomized. Regarding validity, correlation with an internal criterion was better for Likert than for Thurstone-scale, although factorial validity was not satisfactory in neither of them. Results confirm the advantages of Likert-scale and justify the masive use of it. These advantages could be, probably, due to the inclusion in the Likert-scale of items posited in the higher and lower end of the attitudinal continuum.

**Key words:** scale construction, Thurstone scales, Likert scales, reliability, validity.

### 1.- Introducción.

En el ámbito de la psicología se han desarrollado diversas técnicas para medir las actitudes, destacando las propuestas por L. L. Thurstone en 1928 y Remsis Likert en 1932.

**La técnica de THURSTONE** (Thurstone, 1928, 1931) se suele clasificar entre las escalas experimentales, porque los items que se pasan a los sujetos no son fruto de la subjetividad de los investigadores, sino de un proceso experimental previo donde se evalúan objetivamente. Es una técnica elaborada dentro del marco de la **“ley del juicio categórico”** propuesta por Thurstone que sirve para medir cualquier clase de actitud. Esta ley postula que el continuo psicológico puede dividirse en categorías y el sujeto puede asignar directamente valores escalares a los items.

**LIKERT** (Likert, 1932), por su parte, plantea un proceso de elaboración más breve y sencillo que Thurstone y -según el autor- igualmente fiable. Su técnica acepta expresamente que las actitudes pueden medirse a través de manifestaciones verbales y asume la posibilidad de estudiar dimensiones de actitud a partir de un conjunto de enunciados que operen como reactivos para los sujetos. Éstos pueden situarse en la variable de actitud desde el polo más favorable al más desfavorable, de tal manera que la variación de las respuestas será debida a diferencias individuales de los sujetos.

Una breve **comparación** de ambas técnicas nos indica que:

1.- Respecto al método de escalamiento, Thurstone escala fundamentalmente estímulos psicológicos y sobre esta escala sitúa después a los sujetos, mientras que Likert tiene como finalidad principal el escalamiento de sujetos.

2.- Desde el punto de vista de la medición la técnica Likert asume un nivel de medida ordinal. La técnica Thurstone se considera una escala de intervalo, aunque este aspecto es muy cuestionado.

3.- Respecto a los pasos seguidos para la construcción de los dos tipos de escalas:

- Ambas tienen como primer paso definir claramente la dimensión de actitud que se quiere medir e identificar las conductas que reflejen dicha actitud.

- En las escalas Thurstone se pide a los sujetos que emitan sus respuestas con dos alternativas Si/No; mientras que, en Likert se presentan distintas alternativas desde la más favorable a la más desfavorable. Los modelos más habituales varían entre dos y seis respuestas, siendo el número más común el de cinco sugerido por el propio Likert.

- Likert hace la selección de ítems definitivos después de pasar la escala a una muestra representativa de sujetos, Thurstone se basa en los valores asignados por los jueces.

- Las escalas Likert son sumativas y su objeto es el de especificar la valencia (positiva-negativa) de actitud. Las de Thurstone, por su parte, pretenden el establecimiento de un continuo subyacente de la dimensión de la actitud mediante el cálculo de la media de los valores escalares de los ítems a los que el sujeto responde afirmativamente.

Se han señalado (Barbero, 1993; López Pérez, 1989; Morales, 1988) **ventajas e inconvenientes en ambas técnicas**:

En relación con las **escalas Thurstone** se suelen resaltar como **ventaja** su carácter objetivo, porque el valor de los ítems está basado en la prueba de jueces, mientras que como **objeciones** más relevantes se señalan un proceso de elaboración largo y complejo y la posible subjetividad de los jueces. Se considera asimismo que es una escala ordinal más que de intervalo. Además, la unidimensionalidad postulada no es tal porque se introducen otras dimensiones distintas a las que se quiere medir y la estabilidad de los valores escalares puede verse afectada por el número de estímulos y jueces utilizados. Sin embargo, Ponsoda, García y Olea (1989) no obtienen evidencia alguna de mayor estabilidad cuando se emplean más de 30 jueces.

Por lo que se refiere a **Likert**, se señalan como **ventajas** que sus escalas son más

sencillas de contestar, requieren menor trabajo y se realizan de modo más rápido, además de necesitar menor número de ítems para su confección. Parece confirmada la mayor fiabilidad del método Likert, incluso con menor número de enunciados. Como **desventajas**, se le achaca que no permiten determinar en cuánto es más favorable un sujeto que otro respecto a la variable medida. Tampoco es posible establecer en aplicaciones posteriores cuál es la cantidad de cambio experimentado en los sujetos. Por último la puntuación de un sujeto suele tener un significado algo confuso y presenta mayores dudas que la de Thurstone en cuanto a si mide una sola dimensión de actitud.

A partir de las anteriores consideraciones, **nos planteamos el objetivo de intentar comprobar empíricamente las supuestas ventajas teóricas del método Likert.** En la revisión de la literatura sobre el tema, encontramos que algunos autores han tratado de reconciliar los dos acercamientos. Así, Andrich (1989, 1996) plantea un “**modelo de desdoblamiento**” para respuestas directas y lo generaliza después para proporcionar un modelo para respuestas politómicas del estilo Likert. Sin embargo, en cuanto a comparar directamente las dos técnicas, no encontramos ninguna referencia bibliográfica en el período 1991-1997; si bien, son muchos más los trabajos donde se utilizan escalas Likert. Ante tal carencia, nuestro trabajo consistirá en construir la misma escala con ambas técnicas y someterla después a los análisis estadísticos oportunos para determinar empíricamente la supuesta superioridad del método Likert sobre el método Thurstone.

## 2.- Método

En primer lugar definimos el concepto de *reparto del trabajo*. Se podría entender como la distribución de las horas de trabajo anuales totales entre el mayor número de trabajadores posibles con el objetivo de reducir el desempleo al máximo. Sería una modalidad de flexibilización laboral.

Para llevar a cabo nuestros propósitos experimentales se creó una lista de 60 enunciados que reflejasen actitud, tanto positiva como negativa, hacia el reparto del trabajo. De estos 60, en una primera fase se eliminaron 11 debido a su mala formulación.

Siguiendo los criterios de construcción de escalas de Thurstone, se aplicó la lista a un grupo de 37 jueces (34 trabajadores, 2 estudiantes y un jubilado). Éstos tuvieron que puntuar de 1 a 5 el grado en que consideraban que cada enunciado reflejaba actitud favorable (o desfavorable) hacia el reparto del trabajo. Posteriormente volvieron a ser evaluados los enunciados por un grupo distinto de jueces conformado por 26 sujetos (24 trabajadores y 2 estudiantes). La correlación encontrada entre ambas evaluaciones fue  $r = 0.981$  ( $p = 0.000$ ). Utilizando el método de intervalos aparentemente iguales se calculó el *valor escalar* (mediana de las puntuaciones otorgadas por los jueces en cada ítem) y el *índice de ambigüedad* (amplitud intercuartil)

Se establecieron 8 intervalos de  $\frac{1}{2}$  unidad entre 1 y 5, seleccionando un número similar de ítems en cada uno de ellos para construir una escala de 30 ítems, de tal manera que:

- ▶ De los dos intervalos extremos se seleccionaron los enunciados cuyos índices de ambigüedad fueron  $< 1$ .
- ▶ De los centrales aquellos ítems con índices de ambigüedad  $< 2$ .
- ▶ Del resto se estableció como criterio de inclusión que el índice de ambigüedad fuese  $<$

1,5.

Como nuestra intención era aplicar conjuntamente una escala siguiendo la metodología de Likert, a los 30 ítems anteriores se añadieron los 19 restantes, quedando una escala final con 49 enunciados. Por último, se introdujo un ítem en el que se pidió a los sujetos directamente que reflejaran su actitud hacia el reparto del trabajo ("variable 50").

El número de alternativas se estableció en 6 (Muy en Desacuerdo, Desacuerdo, Bastante en Desacuerdo, Bastante de Acuerdo, Acuerdo y Muy de Acuerdo), con el objetivo de poder transformar las respuestas en los ítems correspondientes a la "escala Thurstone" en 0 y 1, y en consonancia con un grupo importante de autores (véase Morales, 1988) que sugieren que este número de alternativas es el óptimo.

La escala completa se aplicó a una muestra de 175 sujetos (79 hombres y 96 mujeres), compuesta por 87 sujetos sin ninguna actividad laboral y 88 en situación de actividad laboral.

### 3.- Análisis de datos

Los análisis estadísticos fueron realizados con el programa SPSS para Windows (versión 7.5). Dependiendo del análisis se utilizaron las siguientes técnicas:

- |                            |                                  |
|----------------------------|----------------------------------|
| a) $\alpha$ de Cronbach.   | d) Prueba de Kolmogorov-Smirnov. |
| b) Correlación de Pearson. | e) Análisis de Cluster           |
| c) Análisis Factorial.     |                                  |

Para obtención de puntuaciones se realizaron las siguientes operaciones:

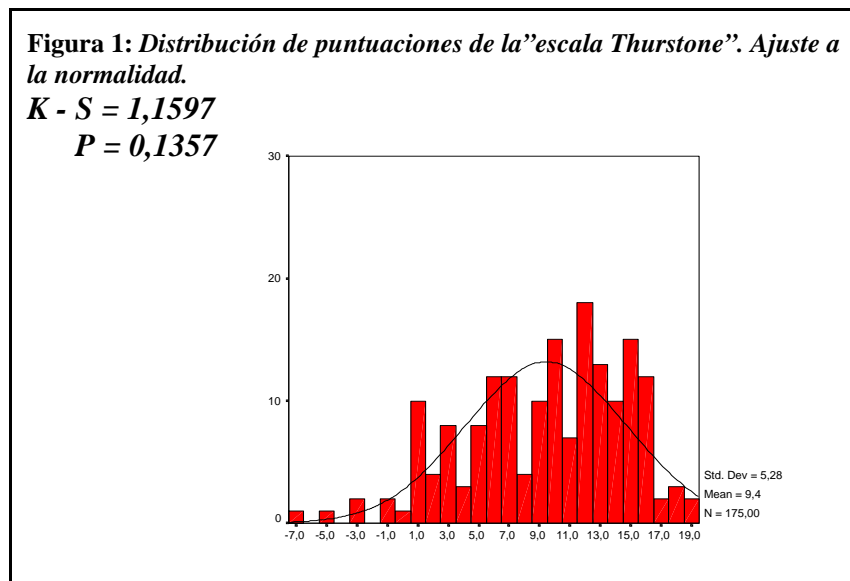
1.- Las puntuaciones de la "escala Thurstone", se recodificaron en dicotómicas 0 para MD, D, BD y 1 para BA, A y MA. A su vez, se realizó una segunda transformación resultante de aplicar la fórmula  $y_j = x_j (VE - 3)$ , siendo  $x_j$  la puntuación de cada ítem y  $VE$  el valor escalar de cada uno de ellos. La puntuación total del test fue  $\sum y_j$ .

2.- Para la "escala Likert", la transformación consistió en invertir las puntuaciones en los enunciados que expresaban actitud negativa. Se procedió a maximizar el coeficiente  $\alpha$ , tras lo cual se obtuvo la puntuación total.

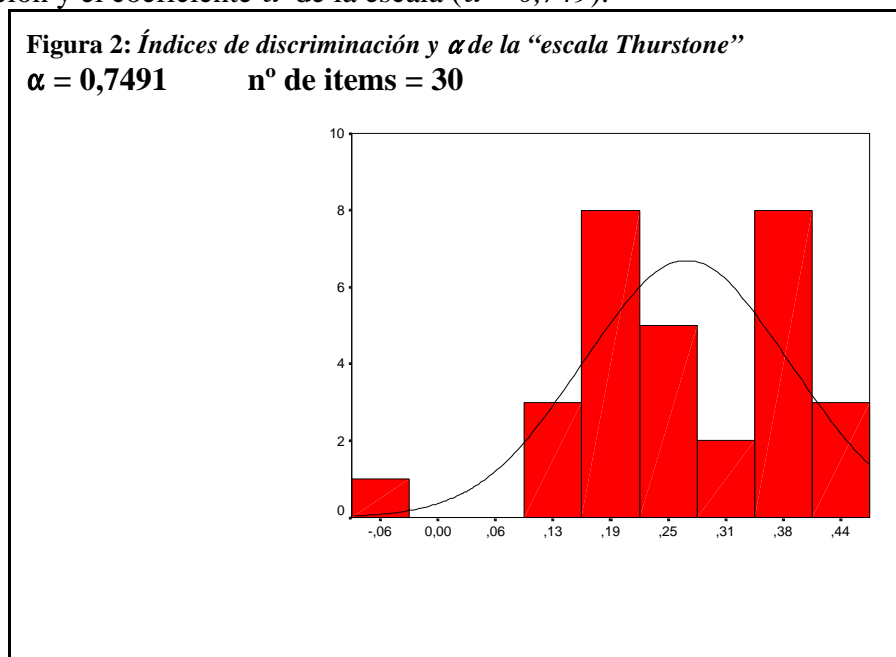
### 4.- Resultados

#### A.- Escala Thurstone:

En la Figura 1, se expresa la distribución de las puntuaciones para el total de la escala, obsérvese que el resultado de la prueba de Kolmogorov-Smirnov fue:  $K-S = 1,1597$  ( $p = 0,1357$ ), por lo que aceptamos la hipótesis de normalidad.



Así mismo, en la Figura 2, se puede apreciar la distribución de los índices de discriminación y el coeficiente  $\alpha$  de la escala ( $\alpha = 0,749$ ).

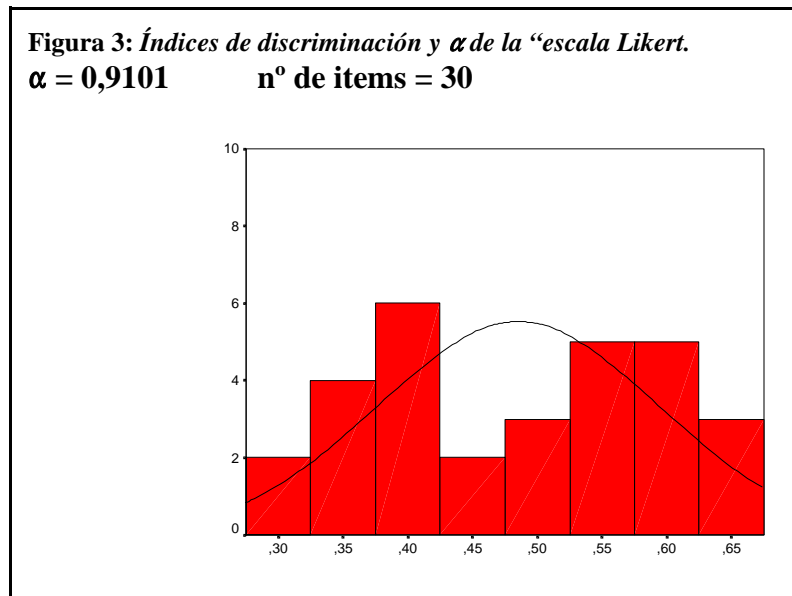


Para comprobar la validez se calculó la correlación entre las puntuaciones totales y la "variable 50" (criterio interno de validez), siendo ésta  $r = 0,543$  ( $p = 0,000$ ).

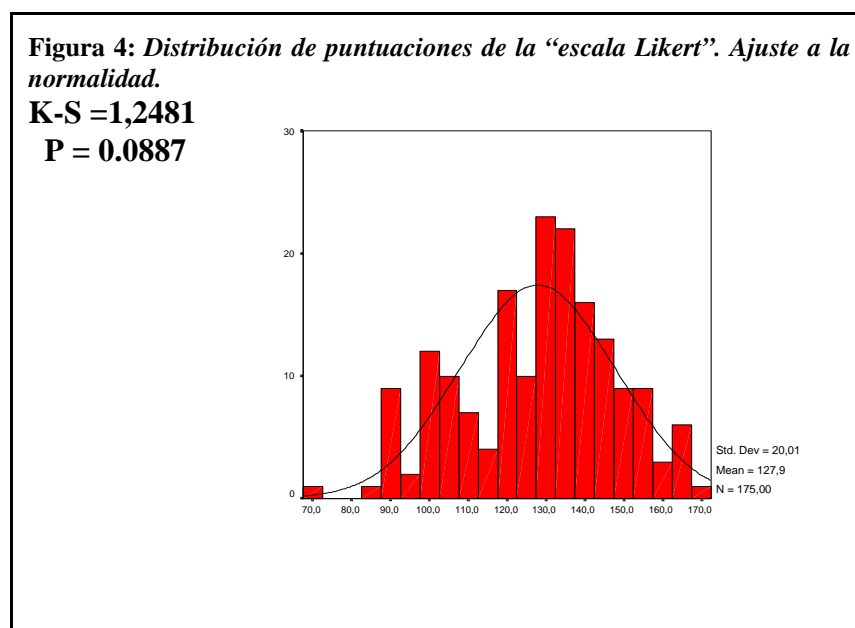
Por lo que respecta a la validez factorial de la escala se realizó un análisis factorial por el método de componentes principales. Siguiendo el criterio de Kaiser-Meyer se obtuvieron 10 factores que explicaban el 62.9 % de la varianza.

### B.- Escala Likert:

En este caso, se procedió a buscar estadísticamente el mayor valor del coeficiente de fiabilidad (ver Figura 3). De esta manera, se obtuvo una escala compuesta por 30 ítems con un  $\alpha = 0,9101$  (con 6 alternativas de respuesta).

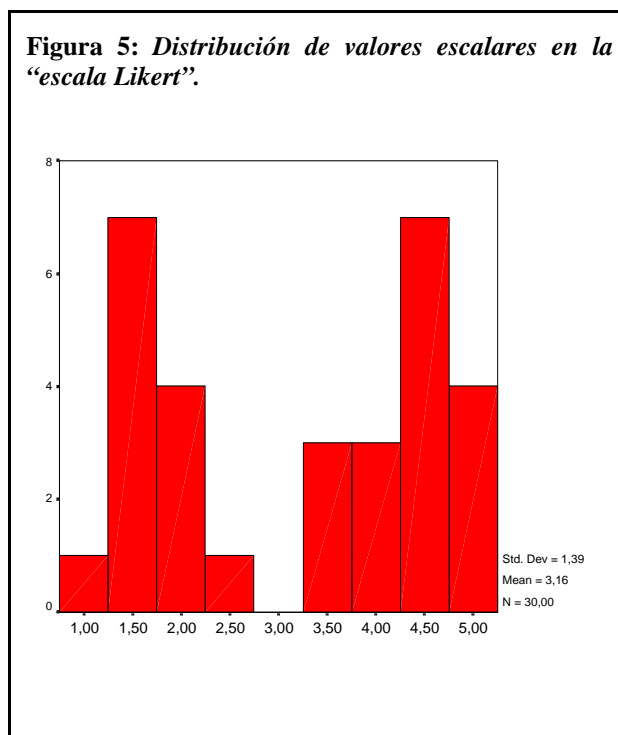


Posteriormente examinó la distribución de puntuaciones (ver Figura 4) y se aplicó la prueba Kolmogorov-Smirnov ( $K-S = 1,2481$ ;  $p = 0,0887$ ). La correlación de las puntuaciones totales con la "variable 50" fue  $r = 0,6993$  ( $p = 0,000$ ) y con la puntuación total de la "escala Thurstone"  $r = 0,806$  ( $p = 0,000$ ).



Con el objetivo de investigar si los ítems se distribuían de una manera uniforme a lo largo del continuo de actitud, se examinó la distribución de los valores escalares (mediana de las

puntuaciones otorgadas por los jueces) de los mismos. Como puede observarse en la Figura 5, los ítems que tuvieron mayores índices de discriminación (aquellos que optimizaron el valor de  $\alpha$ ) se agruparon en torno a los valores escalares extremos.



Por otro lado, nos propusimos estudiar en qué medida se veía afectado el coeficiente de fiabilidad de la "escala Likert" si se reducía el número de alternativas de respuesta. Para ello utilizamos distintas variaciones:

Con 3 alternativas y 30 ítems  $\rightarrow \alpha = 0,8931$ .

Con 2 alternativas y 30 ítems  $\rightarrow \alpha = 0,8650$ .

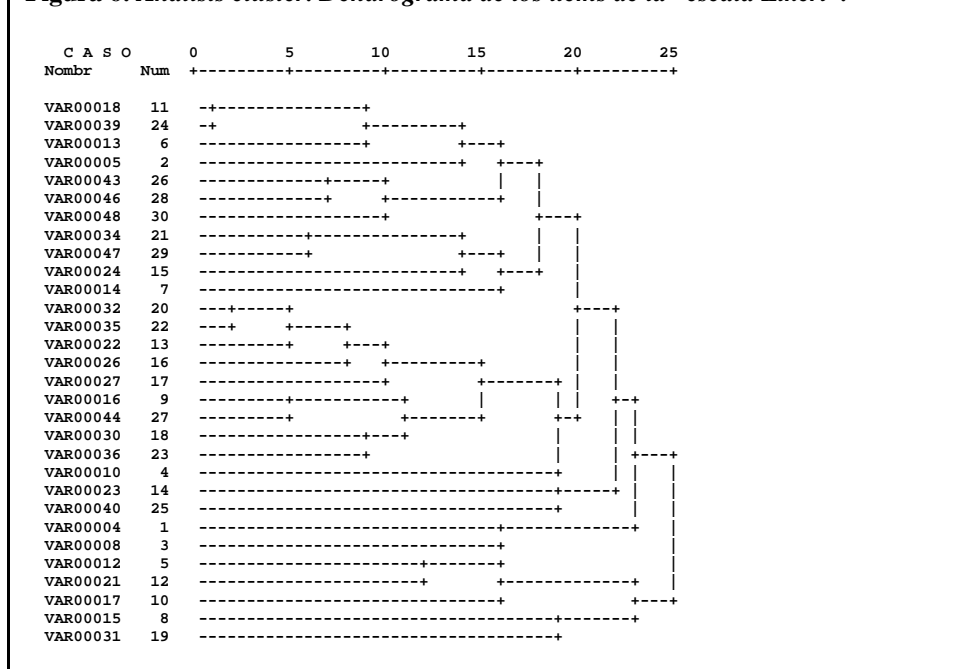
Por último los coeficientes  $\alpha$  de los 18 ítems comunes a ambas escalas fueron:

	Likert	Thurstone
Con 6 alternativas	0,8381	
Con 3 alternativas	0,8126	
Con 2 alternativas	0,7568	0,7503

Con el fin de optimizar la validez factorial se procedió del siguiente modo:

1.- En primer lugar se realizó un análisis de cluster de manera exploratoria, utilizando como método de conglomeración la vinculación inter-grupos y como medida la correlación de Pearson. Se obtuvieron dos clusters, tal y como se muestra en la Figura 6: el primero formado por 11 ítems y el segundo por 9 (total 20 ítems); el resto de los ítems formaron clusters individuales por lo que fueron rechazados.

**Figura 6: Análisis cluster. Dendrograma de los items de la “escala Likert”.**



2.- Posteriormente se procedió a hacer un análisis factorial de estos 20 ítems, utilizando el mismo método de extracción y de retención de factores que se siguió para la “Escala Thurstone”. De esta manera, se obtuvieron 4 factores (véase Tabla I) que explicaban el 58.4 % de la varianza (hay que señalar que con dos factores se explica el 45,6 % de la varianza).

Tabla I: Análisis Factorial, Total Varianza Explicada

Componentes	Autovalores iniciales			Extracción suma de los cuadrados		
	Total	% de Varianza	% Acumulado	Total	% de Varianza	% Acumulado
1	7,348	36,742	36,742	7,348	36,742	36,742
2	1,778	8,890	45,632	1,778	8,890	45,632
3	1,454	7,270	52,902	1,454	7,270	52,902
4	1,102	5,510	58,412	1,102	5,510	58,412
5	,973	4,866	63,278			
6	,836	4,179	67,457			
7	,779	3,893	71,351			
8	,742	3,708	75,059			
9	,689	3,414	78,503			
10	,586	2,930	81,433			
11	,544	2,719	84,152			
12	,527	2,636	86,788			
13	,459	2,295	89,083			
14	,450	2,251	91,334			
15	,381	1,904	93,238			
16	,333	1,665	94,903			
17	,301	1,506	96,409			
18	,274	1,368	97,777			
19	,247	1,237	99,014			
20	,197	,986	100,000			

Método de Extracción: Componentes Principales

En los dos primeros factores (ver Tabla II) cargaron todos los ítems relacionados con una



dimensión cognitivo-evaluativa: el primero relacionado con regulación “normativa” del reparto del trabajo y el segundo, con las consideraciones sobre si mejora o empeora la calidad del trabajo en caso de repartirse el mismo. Los factores 3 y 4 tienen que ver con las implicaciones personales (planteamientos económicos, solidarios...).

Tabla II: Matriz de Componentes Rotados

	Componentes			
	1	2	3	4
VAR00005			.602	
VAR00013			.631	
VAR00014		.646		
VAR00016				
VAR00018			.704	
VAR00022	.766			
VAR00024		.511		
VAR00026	.737			
VAR00027	.604			
VAR00030		.685		
VAR00032	.701			
VAR00034		.633		
VAR00035	.760			
VAR00036		.592		
VAR00039			.753	
VAR00043				.503
VAR00044				
VAR00046				.748
VAR00047		.572		.549
VAR00048				

Método de Extracción: Componentes Principales.

Se calculó la fiabilidad de estos 20 ítems, siendo  $\alpha = 0,9055$ . También se obtuvieron las puntuaciones totales y se aplicó la prueba de Kolmogorov-Smirnov ( $K-S = 1,1646$ ;  $p = 0,1327$ ). Las correlaciones que se encontraron con la “variable 50” y con las puntuaciones de la “escala Thurstone” fueron  $r = 0,6674$  ( $p = 0,000$ ) y  $r = 0,757$  ( $p = 0,000$ ), respectivamente.

## 5.- Conclusiones

Los resultados del presente trabajo permiten comprobar empíricamente las supuestas ventajas teóricas de la técnica Likert de construcción de escalas frente a la técnica de Thurstone. El procedimiento utilizado permite la comparación de ambas técnicas, puesto que se partió de la misma muestra e idéntica base de ítems, que fueron seleccionados y puntuados conforme a las directrices características de ambos tipos de escalas.

Las ventajas más claras a favor de la técnica Likert se dieron en relación con la fiabilidad. El coeficiente  $\alpha$  de la “escala Likert” alcanzó un valor de 0,910 mientras que el de la “escala Thurstone” fue de 0,749. Esta diferencia no se debe al número de ítems incluidos en cada escala, dado que fue el mismo en ambos casos; incluso al calcular  $\alpha$  para la escala Likert utilizando sólo los 18 ítems comunes con la escala Thurstone se mantuvo la ventaja ( $\alpha = 0,838$ ). Cabría esperar que las diferencias entre ambas escalas se puedan deber a: 1) la forma de seleccionar los ítems, 2) el formato de respuesta (número de alternativas) y/o 3) el modo de asignar puntuaciones.

No parece probable que la mayor fiabilidad de la escala Likert se deba al mayor número de alternativas de respuesta. Confirmando lo apuntado por diversos autores (p.e. Chang, 1994) el coeficiente de fiabilidad de dicha escala varió en función del número de alternativas, sin

embargo, aun con dos alternativas continuó siendo superior ( $\alpha = 0,865$ ) al de la escala Thurstone.

Las diferencias encontradas podrían deberse a los otros dos factores señalados. En lo que respecta a la forma de seleccionar los items, en las escalas Likert cuando se localizaron los items en un continuo según los principios de Thurstone se situaron en los extremos con un hueco en el medio (véase figura 5), cuestión que contrasta con los requisitos de Thurstone, donde los enunciados son localizados, más o menos, uniformemente en el continuo. Esta puede ser una de las causas de la mayor fiabilidad de la escala Likert, tal y como plantea Andrich (1989). Pero incluso eliminando las diferencias producidas por el método de selección de items y el número de alternativas (es decir, con los 18 items comunes las dos escalas), la fiabilidad de la escala Likert fue similar a la de Thurstone. Así pues, la asignación de puntuaciones según la metodología propuesta por Thurstone no parece aportar ninguna ventaja frente a una puntuación dicotómica de la escala Likert. Dada esta situación podría pensarse que los valores escalares asignados no son adecuados a consecuencia de la subjetividad de los juicios emitidos por los jueces; sin embargo esto quedaría descartado por el alto coeficiente de acuerdo inter-jueces encontrado ( $r = 0,981$ ).

En relación con la validez utilizamos dos tipos de validación: de criterio y factorial. La validez de criterio fue superior para la escala Likert ( $r = 0,699$ , frente a  $r = 0,543$  para Thurstone). En cuanto a la validez factorial no se alcanzaron resultados satisfactorios en ninguna de las escalas. Sin embargo en la escala Likert se consiguió optimizar la validez factorial (véase Figura 6 y Tablas I y II) sin un menoscabo significativo de la fiabilidad ( $\alpha = 0,905$  con 20 items).

En conclusión nuestros resultados parecen sugerir que uno de los mayores problemas de la técnica Thurstone está en la estrategia seguida para la selección de los items al intentar cubrir de manera uniforme todo el continuo de actitud. En los valores escalares medios cabrían tanto items realmente neutrales como items no directamente relacionados con la variable de actitud que se desea medir. Esto tendría consecuencias negativas en la fiabilidad y validez de la escala que, unidas a su mayor complejidad procedimental, parecen justificar la mayor popularidad de las escalas Likert.

## 6.- Referencias

- Andrich, D. (1989). The application of an unfolding model of the PIRT type for the measurement of attitude. **Applied Psychological Measurement**, **12**, 33-51.
- Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. **British Journal of Mathematical and Statistical Psychology**, **49**, 347-365.
- Barbero, M.I. (1993). **Métodos de Elaboración de Escalas**. Madrid. UNED.
- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. **Applied psychological measurement**, **18** (3), 205-215.
- Likert, R. (1932). A Technique for the measurement of attitudes. **Archives of Psychology**, **140**, 1-55.
- López Pérez, J. (1989). Las escalas de actitudes (2). En Morales (Comp.) **Metodología y Teoría de la Psicología**. Madrid. UNED.
- Morales, P. (1988). **Medición de actitudes en psicología y educación**. San Sebastián: Tarttalo.
- Ponsoda, V., García, C. & Olea, J. (1989). Influencia del número de jueces en la asignación de valores escalares: Un estudio de simulación. **Psicológica**, **10**, 153-161.

- Thurstone, L. L. (1928). Attitudes can be measured. **American Journal of Sociology**, **33**, 529-544.
- Thurstone, L. L. & Chave, E. J. (1929). **The Measurement of Attitudes**. Chicago. University of Chicago Press.
- Thurstone, L. L. (1931). The measurement of social attitudes. **Journal of Abnormal and Social Psychology**, **26**, 249-269.