

CREENCIAS BÁSICAS EN LA CONSTRUCCIÓN DE ESCALAS Análisis Intencional y Procedimiento de Evaluación

Adela Garzón

Adela Garzón es Catedrática de Psicología Social en la Facultad de Psicología (Avda. Blasco Ibáñez 21, 46010-Valencia) de la Universidad de Valencia.

Planteamiento

Es ingenuo pensar que el autor que pretende medir un fenómeno psíquico y se intro-

duce en la complicada tarea de encontrar un sistema de medida, no tiene ya en mente unas expectativas, una idea más o menos consciente de lo que es y las características que rodean al fenómeno de su estudio. Dicho en términos psicológicos, los investigadores poseen un conjunto de creencias e ideas preconcebidas sobre lo que están investigando que, intencionalmente o no, influye en los sistemas de medida que va a construir y en la interpretación de los resultados que obtenga. Teoría de la medida, construcción de escalas de medida, cuantificación e inferencia, elaboración de test, fiabilidad, validez, muestreo y tantos otros términos son habituales dentro de la psicología; se relacionan con las dificultades de utilizar los números para detectar, medir, analizar y predecir fenómenos psíquicos. Todos tienen características específicas en función de las propiedades de los fenómenos a los que se aplican y del tipo de medición que se persigue (Anastasi, 1977, Seoane, 1988) y en todos influye la idiosincrasia del investigador.

El desarrollo de técnicas de investigación y evaluación de distintos aspectos del comportamiento humano es uno de los campos de la psicología que más ha evolucionado y, quizá, más atención y cuidado ha recibido por parte de los especialistas en el campo de la medida psicológica. Su evolución y afinamiento se ha servido y fundamentado en los avances en la elaboración y tratamiento estadísticos de datos, sean cuantitativos o cualitativos, experimentales o de campo, pero la medida psicológica sigue teniendo los mismos problemas métricos de siempre:

primero, la identificación y definición de un aspecto del comportamiento que se quiere evaluar; segundo, la elaboración y validación de una prueba para medir dicho aspecto y tercero la elaboración de los datos brutos, proporcionados por los sujetos y su interpretación. Además, la investigación psicológica puede haber avanzado en técnicas, pero sigue teniendo investigadores con ideas previas sobre lo que investiga.

En la actualidad tendemos a olvidar los primeros pasos (la selección de ítems) sobre todo a la hora de generalizar los resultados obtenidos; equivocamos el sistema de medida del fenómeno investigado con el fenómeno mismo. Gran parte de las disparidades encontradas en investigaciones del mismo fenómeno son resultado de estas confusiones básicas. Dicho de otro modo, damos por sentado o aceptamos con demasiada facilidad el producto acabado, la escala construida, prestando poca atención al proceso de su elaboración, cuando interpretamos datos.

¿Dónde radica nuestra innovación en la cuestión de selección y análisis de ítems? Dicho de forma rápida, en la presentación de un procedimiento para un análisis interno de la propia escala, sin necesidad de utilizar muestras de sujetos, o más exactamente, antes de que se pretenda validar a través de muestras de sujetos reales. Teniendo en cuenta una valoración y análisis interno, autóctono (sin intermediarios) se puede determinar qué se está midiendo, qué "construcción social" del tema de investigación está implícita y, en consecuencia, hasta qué punto se puede generalizar en términos conceptualmente la escala construida.

Pongamos un ejemplo; la escala F recibió amplias críticas por estar sobrecargada de preguntas con un contenido claramente conservador. Ahora, casi 60 años después de su publicación, cabe preguntarse si realmente estaba sesgada, o si los autores de la escala F, pretendían medir el autoritarismo que estaban experimentando, el construido al amparo de una ideología concreta y de las coordenadas de una sociedad occidental, en pleno debate ideológico.

El material del análisis

Desde los años 80 hemos utilizado el análisis de contenido como técnica de investigación, pero lo aplicamos a temas concretos de la realidad social. Por ejemplo, a través de la realización de tesis de licenciatura y de doctorado dirigidas por J. Seoane el análisis de contenido sirvió para investigar los valores e ideología de líderes políticos (Moreno, 1985, 1987, en análisis de líderes políticos concretos; Garcés, 1987, en análisis de discursos de presidentes autonómicos), o para analizar las agendas de los medios de comunicación (García de las Bayonas, 1983; Sanfeliu, 1986, 1988) o, bajo la dirección de Garzón, las claves psicológicas del pensamiento militar (Herrera, 1986, Garzón, 1988). Posteriormente procedimientos concretos de análisis de contenido nos fueron útiles para entender las claves del pensamiento y motivos de distintos líderes na-

cionales y autonómicos, análisis que se divulgaron en medios de comunicación.

Ahora, afinando el clásico análisis de contenido (Krippendorff, 1967, 1980, 1986; Strauss, 1987) con los avances tecnológicos, pretendemos desarrollar un procedimiento estandarizado que permita clarificar conceptualmente las escalas de medida que se utilizan en la investigación psicológica.

Para presentar el nuevo procedimiento de análisis interno de las escalas, vamos a utilizar dos que miden el autoritarismo. Nos referimos a la escala de Fascismo (F), elaborada por T. Adorno y colaboradores en 1950 y, por otro lado, la escala del Autoritarismo de Derechas (RWA) de B. Altemeyer, en una de sus últimas versiones, la de 1986 (Altemeyer, 1981, 1988).

Varias razones nos han llevado a seleccionar estas dos escalas. Ambas tratan la misma problemática, tienen una extensión similar en cuanto al número de preguntas que contienen, además disponemos de versiones adaptadas al castellano. Pero la razón más relevante es que tratan una problemática concreta, que indudablemente simplifica la posibilidad de ratificar el procedimiento, y que existe suficiente evidencia empírica de los componentes centrales que definen el tema.

Las versiones de las dos escalas utilizadas para el análisis son, por un lado, la versión de la escala F adaptada por Pinillos (1963) y reducida a 28 ítems. La versión de la RWA es una adaptación realizada por Seoane y Garzón. Aunque Altemeyer concibe su escala de autoritarismo como una escala unidimensional, en la adaptación castellana se realizó una estructura factorial de seis componentes. La descripción y factorización de ambas escalas pueden consultarse en Seoane-Garzón (1992) y en Garzón-Seoane (1996).

Procedimiento

El procedimiento que proponemos es un análisis de contenido de las propias escalas, no tanto para seleccionar y pulir las preguntas, sino para detectar creencias e ideas previas que las fundamentan. La concepción implícita debería ser utilizada en los procesos posteriores de fiabilidad y validez de las escalas.

El análisis de contenido que proponemos está orientado por una directriz: utilizar un procedimiento lo más estricto posible, de forma que las inferencias o decisiones discrecionales del investigador sean las mínimas posibles y seguir la máxima de que "cualquier otro investigador pueda repetir y llegar a los mismos resultados".

Tres partes componen el procedimiento que proponemos: 1) análisis formal de las escalas, 2) análisis de contexto; es decir, características del léxico y expresión, y 3) componentes básicos; es decir, la identificación de las ideas y creencias básicas internas de la escala.

1) Los indicadores utilizados para los aspectos formales de las escalas son dos:

Indicadores formales: número total de palabras (NPT); número de palabras distintas (NPD); promedio de repetición de las palabras de la escala (PRP) y su equivalente índice de redundancia (IR, en una escala de 0 a 1).

Indicadores del léxico y de expresión: promedio de palabras por pregunta; número de preguntas en que aparecen las palabras; número y frecuencia de relaciones de los términos (las veces que un término aparece junto a otro en la misma pregunta), que convertimos en un índice de convergencia que nos sirve para el análisis del contexto.

En este punto debemos señalar que los indicadores generales elegidos son los que se adecuan al tipo de material con que trabajamos. Es decir, estamos tratando con un material restringido —formulación de preguntas cerradas y delimitadas en su extensión. Disponemos de otros indicadores léxicos y de expresión cuando el material-base es un texto abierto, que entonces permite utilizar más componentes o indicadores formales y medidas comparativas internas al propio contenido. Sin embargo, en este trabajo no tiene sentido hablar de ellos.

2) Para el segundo paso, denominado *Análisis de contexto*, es necesario que previamente se determine la unidad elemental de análisis. En el caso que nos ocupa son los términos; dependiendo del material y de los objetivos se pueden establecer unidades de análisis más amplias.

Las *unidades de análisis* o selección de las unidades relevantes en las que se basa el análisis de contexto, servirán para detectar los contenidos centrales de la escala.

Cuando se realiza un análisis interno de contenido y se miden frecuencias de las unidades de análisis, es necesario tomar dos decisiones previas: la primera es cuál será la unidad de análisis más elemental. La segunda es el criterio de selección de dichas unidades.

En el procedimiento que proponemos y para evitar inferencias o juicios previos de los investigadores, introduciendo así variables extrañas, proponemos que la unidad de análisis elemental sean “los términos”. En cuanto al criterio de selección, optamos por una decisión de tipo cuantitativo: primero seleccionar las palabras de mayor frecuencia y de estas recoger las que tienen un significado por sí mismas. Es decir, se descartan: artículos definidos e indefinidos, adjetivos, adverbios, pronombres y todo término que no contenga en sí mismo un significado relevante para la investigación.

Para hallar la frecuencia de un término en la escala se suman las variantes de dicho término en función de que tengan la misma raíz. Un ejemplo; la frecuencia del término “débil”, será la suma de frecuencias de todos los términos con la misma raíz (por tanto, se suman las fre-

cuencias de los términos “débil, débiles, debilitamiento, debilidad, debilidades”, etc.).

Existen criterios más amplios y quizá inicialmente más potentes para el análisis interno, pero éstos introducen inevitablemente interpretaciones subjetivas de los investigadores. Para evitar tales inferencias discrecionales, hemos optado por un criterio más restrictivo, pero en contrapartida, más delimitado y preciso.

El *análisis del contexto* proporciona una matriz de asociación, indicando qué términos se asocian entre sí y cuáles no. Esta matriz sería similar a una retícula semántica, utilizando la terminología del Procesamiento de Información (Lindsay y Norman, 1975; Seoane, 1979; Garzón-1992) que revela parte de los elementos del sistema de creencias que pone en juego el investigador al construir la escala.

3) El análisis de los componentes del contenido de la escala. El tercer paso, una vez seleccionadas dichas unidades y analizada la matriz de contexto, será detectar los componentes centrales que definen el conjunto de ideas previas utilizadas por el investigador. Estos componentes lógicamente no pueden ser los mismos que la validación empírica de la escala a través de sujetos, aunque deberían detectar, si nuestro procedimiento es acertado, algunas de las claves de dicha validación empírica posterior.

Algunas creencias básicas de dos escalas de Autoritarismo

Empezaremos por describir el análisis desarrollado por la escala de Autoritarismo más recientemente publicada, la escala RWA de Altemeyer (1988) y después describiremos los resultados obtenidos con la escala F de Adorno y el grupo de Berkeley (1950).

La escala RWA

Aspectos formales

En las Tabla 1 y 3 se presentan los indicadores formales y los relacionados con el léxico y expresión. Las partes o estructura interna del texto que utilizamos para analizar los ítems de la escala se corresponden en este caso con el número de preguntas que forman la escala. Es decir, dividimos el texto en 30 partes correspondientes a los 30 ítems. Solamente en un ítem, aparecen dos frases, el resto presenta una única frase. Optamos, dado que era posible gramaticalmente, unir ambas frases, por la única razón de simplificar el análisis y dado que solamente un caso presentaba dicha característica. Por eso, las frases y las partes, coinciden en este caso.

El total de términos utilizados en la RWA es de 700, de los que 322 son diferentes. Cada término aparece en un promedio de algo más de dos veces (2,17 veces más).

Este promedio, es equivalente a un Índice de Redundancia, obtenido a partir de la división del número total de las distintas por el número total de palabras. Este índice es una escala estandarizada de 0 a 1, en la dirección de poca redundancia. En el caso de la RWA el índice de redundancia no llega a la media, aunque se acerca (0,46). Por último, el número medio de términos utilizados en cada frase es de 23,33.

Tabla 1
Escala RWA
Indicadores formales, de léxico y expresión

Segmentos	30
Frases	30
Total de términos	700
Nº términos distintos	322
Promedio de repetición	2,17
Índice de redundancia	0,46
Nº de Párrafos	30
Media palabras por sentencia	23,33

Del total de términos que forman la escala RWA, las unidades de análisis que hemos seleccionado para nuestra investigación son aquellos términos que, una vez sumados sus derivados (recordemos que la suma de frecuencia se hace en función de la raíz de las palabras), obtienen una frecuencia superior a 4 y tienen significado en sí mismos.

Estos términos quedan reducidos, en el caso que nos ocupa, a 11 términos que son: autoridad, costumbres, deber, gente, moral, orden, padre, país, personas, sexual y sociedad. Sus frecuencias totales de aparición en la escala es la siguiente:

Tabla 2
Escala de Altemeyer
Términos con frecuencia superior a 2 seleccionados

<i>Términos</i>	<i>frecuencia</i>	<i>items donde aparecen</i>
autoridad	9	3, 5, 12, 14, 22, 23, 27, 29, 30
costumbres	4	13, 24, 29, 30
deber	9	4, 9, 12, 13, 14, 17, 19, 22, 29
gente	4	4, 15, 23, 30
moral	5	4, 9, 10
orden	4	9, 17, 17, 20
padre	5	7, 22, 23, 29, 30
país	4	1, 14, 16, 28
personas	5	6, 10, 18, 24, 27
sexo	4	8, 9, 21, 24
sociedad	4	3, 14, 29, 30

Solamente las preguntas 2, 11, 25 y 26 no recogen ninguno de los términos seleccionados.

Aspectos del léxico y de expresión

En la tabla 3 presentamos una matriz que refleja las relaciones de contexto entre los términos. Se recoge en esta matriz el número de veces que un término aparece junto a otro en la misma frase. La matriz completa del contexto nos proporciona una imagen, una especie de red semántica, por utilizar términos del clásico Procesamiento de Información (Seoane, 1979), o en términos de la psicología social una especie de sociometría, aplicada al sistema de creencias del autor.

En términos psicológicos esta matriz contextual nos permite, al menos, intuir algunos elementos centrales del sistema de creencias del autor. En la matriz de contexto existen tres tipos de información: el número de asociación de términos, la proporción de repeticiones reales frente a las posibles (índice de convergencia) y el total de repetición de las asociaciones encontradas (índice de repetición de asociaciones).

En la escala RWA se dan 25 asociaciones de términos de las 55 posibles. El índice de convergencia es de .45, (proporción entre las asociaciones observadas y las teóricas o posibles, de 0 a 1 en la dirección de mayor convergencia. Las repeticiones de dichas asociaciones es 50 (suma total, ver tabla 3).

Tabla 3

La matriz de Contexto interno de los términos de la escala RWA

términos	autoridad	costumbre	deber	gente	moral	orden	padre	pais	personas	sexo	sociedad	total
autoridad	9	4	8	4			8	4			8	6
costumbres		4	4	2			4		2	2	4	7
deber			9	2	4	2	4	2			4	8
gente				4	2		4				2	6
moral					5				2			3
orden						4						1
padre							5				4	5
país								4			2	3
personas									5	2		3
sexo										4		2
sociedad											4	6

En la matriz de contexto del léxico en la RWA existen tres frecuencias recurrentes: dos, cuatro y ocho son las veces en que un término va junto a otro en la misma frase. Atendiendo a los más recurrentes, vemos que la frecuencia ocho veces se da en los términos “deber, padre y sociedad”; los tres aparecen ocho veces relacionados con el término “autoridad”. Sociedad, al mismo tiempo, aparece cuatro veces con gente y país. Costumbre aparece cuatro veces con deber, padre y sociedad. También “deber” se asocia cuatro veces con moral, padre y sociedad.

En resumen, autoridad, costumbres, deber, gente y sociedad son los términos más relacionados con los otros términos (entre 6 y 8) y los menos, son orden y sexo. Recordamos aquí que hablamos de relaciones entre términos, no de la frecuencia total de relaciones (ésta puede ser muy alta, pero estar limitada a un solo término, o a la inversa).

También es muy revelador atender a los términos que el autor nunca pone juntos en la misma pregunta. Utilizando las ideas de Rokeach (1960), estaríamos hablando, del sistema de descreencias. Desde este otro punto de vista destacamos que en el autor de la escala, el término “moral” no se relaciona con conceptos sociales, sino con lo personal. Cuando menciona lo moral no lo asocia a autoridad, costumbres, orden, país, sociedad, sino a persona, gente y deber. El orden, por otro lado, es el término que se relaciona con un solo término, el “deber”.

El análisis posterior de los componentes centrales de éstos términos nos permitirá simplificar esta telaraña de relaciones contextuales encontrada en la escala RWA.

Análisis de los componentes básicos

Aunque la matriz del contexto de los términos nos acerca más que la simple frecuencia de términos a la forma que el autor tiene de formular el problema que mide en su escala, el análisis más completo de sus ideas sobre el tema, al margen de la validación a través de muestras de sujetos, es precisamente el análisis factorial.

El procedimiento utilizado para hacer el análisis factorial fue utilizar como sujetos las 30 preguntas de la escala y como variables los once términos con los que trabajamos desde el principio. El resultado se expone en la Tabla 4.

La explicación de la varianza de cada variable por el conjunto de los componentes extraídos (comunalidades) no es muy alta. Los términos menos explicados son gente y sexo (,285 y ,316 respectivamente) y los más altos: autoridad, deber y padre, que se sitúan por encima de ,60).

La varianza explicada, en la matriz rotada, con cuatro factores es de 64%. Se utilizó el método de extracción de componentes principales, con rotación Varimax y en la extracción de número de factores, se utilizó el criterio de explicación de 1 de varianza. En la segunda opción de rotación obtuvimos tres factores, que en este caso explica el 53,03 y que al ser más simple y clara optamos por esta segunda rotación (Tabla 4).

Tabla 4.
Análisis factorial de términos seleccionados de la RWA*
Varianza total explicada

Com	Autovalores iniciales			Suma saturaciones al cuadrado de la rotación		
	Total	% varianza	% acum.	Total	% varianza	% acum.
1	2,674	24,312	24,312	2,261	20,557	20,557
2	1,673	15,207	39,520	1,968	17,893	38,450
3	1,487	13,514	53,034	1,604	14,584	53,034
4	1,199	10,903	63,937			
5	,925	8,413	72,350			
6	,822	7,472	79,821			
7	,616	5,600	85,421			
8	,586	5,325	90,747			
9	,474	4,305	95,051			
10	,324	2,945	97,996			
11	,220	2,004	100,000			

*Matriz de correlaciones (KMO=.502; prueba de Bartlett: chi-cuadrado 63,78, gl. 55, sig.=,19).

Tabla 4 (continuación)
Matriz de componentes rotados(a)

	Componente		
	1	2	3
padre	,765		
costumbres	,705		
sociedad	,609	-,429	
gente	,528		
país		-,655	
autoridad	,542	-,625	
sexual		,549	
moral		,517	,463
deber			,773
orden			,671
personas		,502	-,545
Varianza exp.	2,261	1,968	1,604
%	20,557	17,893	14,584

Método de extracción: Análisis de componentes principales. Rotación Varimax.

En la presentación de la matriz factorial rotada omitimos las cargas factoriales inferiores a .30 para facilitar una comprensión rápida de los componentes.

La matriz de componentes rotados nos proporciona algunas claves conceptuales que el autor de la RWA maneja en su escala de autoritarismo. El primer conjunto revela un factor de *convencionalismo*, definido por el papel de la figura tradicional del padre y las costumbres en relación con la sociedad y la gente.

El segundo componente está definido de forma bipolar, pero con un sentido claro en su problemática. Se pone en juego la contraposición entre la autoridad y país, frente a cuestiones de elección personal (lo moral, inmoral y la sexualidad). Es la cuestión de una visión de lo moral y lo sexual en oposición a las directrices de la autoridad y lo que conviene a un país. En este sentido y atendiendo a los ítems en que los términos aparecen, es un componente que guarda relación con algunos aspectos contenidos en la idea de la agresión y sumisión autoritaria, pero desde ese eje de contraponer lo institucional frente a lo personal, haciendo hincapié en los aspectos de moral y sexualidad. Entendemos que este componente alude a una confrontación de la *Autoridad y la desviación moral*. El tercer componente, es un factor de agresión autoritaria en el que el deber y el orden se contraponen a las personas.

La escala F

Aspectos formales

En las Tablas 5, 6 y 7 se presentan los indicadores formales y los relacionados con el léxico y expresión de la escala F de Adorno.

El total de términos utilizados es de 430, de los que 242 son diferentes. Cada término no llega a repetirse dos veces, en un promedio (1,77). El índice de redundancia es de .56 es decir, poco redundante (recordemos que el índice se mueve entre "0" y "1" en la dirección de poca redundancia). Por último, el número medio de términos utilizados en cada frase es de 15,36.

Como en la escala RWA, el criterio de estructuración interna del texto fue dividir el texto por el número de ítems. En este caso son 28 partes y cada parte consta de una sola frase, es decir, 28 frases.

Del total de términos que forman la escala F y siguiendo el criterio general que indicamos en el procedimiento, se seleccionaron para el análisis de contenido aquellos términos que una vez sumadas su raíces y derivados, obtuvieran una frecuencia, en este caso, igual o superior a 2. Inicialmente seleccionamos los de frecuencia igual o superior a tres, pero dada la mayor variación (promedio de repetición e índice de redundancia) mostrada por la escala F (ver tabla 5) decidimos rebajar el criterio a 2 de frecuencia.

Tabla 5

Escala F (versión J.L.Pinillos)
Indicadores formales, de léxico y expresión

Segmentos	28
Frases	28
Total de términos	430
Nº términos distintos	242
Promedio de repetición	1,77
Índice de redundancia	0,46
Nº de Párrafos	28
Media palabras por sentencia	15,36

Estos términos en el caso de la Escala F son 12 términos: ciencia, deber, evitar, gente, guerra, ideas, jóvenes, niños, obedecer, persona, respeto y sexo. Sus frecuencias totales de aparición en la escala es la siguiente:

Tabla 6

Escala F

Términos con frecuencia superior a 2 seleccionados

<i>términos</i>	<i>frecuencia</i>	<i>Items en que aparecen</i>
ciencia	3	3, 17, 18
deber	5	1, 7, 18, 24, 27
evitar	3	7, 10, 27
gente	4	7, 15, 23, 27
guerra	2	4, 19
idea	2	11, 26
jóvenes	2	8, 11
niños	2	1, 21
obedecer	2	1, 5
persona	5	9, 16, 18, 23, 26
respeto	3	1, 16, 27
sexo	3	14, 21, 24

Las preguntas 2, 6, 12, 13, 20, 22, 25 y 28 de la escala F no recogen ninguno de los términos seleccionados.

Aspectos del léxico y de expresión

En la tabla 7 presentamos una matriz que refleja las relaciones de contexto entre los términos seleccionados. Es decir, las veces que dos términos aparecen juntos en la misma frase o ítem.

En la matriz de contexto del léxico en la escala F vemos una tendencia a relacionar los términos pocas veces. De hecho las relaciones de contexto que dominan son mayoritariamente dos veces, solamente esta frecuencia alcanza cuatro repeticiones en deber, evitar, gente y respeto. Un perfil que nos lleva a pensar que una de las claves de la escala F, en función de los términos más frecuentes utilizados y relacionados, es el de *prejuicio*, en un sentido genérico de desprecio y distanciamiento “como obligación de la gente”. Deber, respeto, persona, niños, y gente son los términos, por ese orden, que más asociaciones presentan.

También es muy revelador atender a los términos que el autor nunca pone juntos en la misma pregunta. Es el caso de “jóvenes” que se aparta de todos los términos, excepto del “idea”. Por contraposición deber y respeto son los más interrelacionados con el resto. Tres claves apuntamos: prejuicio, conformismo (joven e idea), y valores tradicionales (deber y respeto)

Tabla 7
La matriz de Contexto interno de los términos de la escala F

	ciencia	deber	evitar	gente	guerra	idea	jóvenes	niños	obedecer	persona	respeto	sexo	total
ciencia	3	2								2			2
deber		5	4	4				2	2	2	4	2	8
evitar			3	4							2		3
gente				4						2	2		4
guerra					2								1
idea						2	2			2			2
jóvenes							2						1
niños								2	2		2	2	4
obedecer									2		2		3
persona										5	2		5
respeto											3		6
sexo												3	2

Lo más significativo, en comparación con la escala RWA, a nivel de estilo de expresión y teniendo en cuenta que hablamos no de las escalas, sino de la selección que hemos realizado de términos más frecuentes, es que en la escala F los términos asociados están más definidos en conjuntos conceptuales (existen 66 relaciones posibles y 20 reales). Su índice de convergencia es de ,30 (recordemos que este índice va de 0 a 1 en dirección de mayor convergencia). También las asociaciones de términos tienen menor frecuencia de repetición y son menos (41, ver columna *total* en la Tabla 7). En la escala RWA los términos tienen mayor número de asociaciones (más contexto de los términos), aunque la frecuencia es mayor. Se repiten más dichas asociaciones. Hay menos diferenciación o dicho de otra manera, mayor convergencia.

Análisis de los componentes básicos

Con la finalidad de detectar componentes conceptuales que utiliza la escala F para formular las preguntas, realizamos también un análisis factorial de la matriz de correlaciones. Como en el caso anterior, los sujetos del análisis factorial son las 28 preguntas de la escala y las variables los términos de mayor frecuencia en el conjunto de la misma. Los resultados se exponen en la Tabla 8.

Tabla 8
Análisis factorial de términos seleccionados de la Escala F*
Varianza total explicada

Comp.	Autovalores iniciales			Suma de las saturaciones al cuadrado de la rotación		
	Total	% varianza	% acu.	Total	% varianza	% acu.
1	2,384	19,868	19,868	2,025	16,873	16,873
2	1,834	15,287	35,155	2,019	16,823	33,696
3	1,488	12,398	47,553	1,563	13,022	46,718
4	1,356	11,297	58,850	1,456	12,132	58,850
5	1,175	9,795	68,645			
6	,973	8,107	76,752			
7	,807	6,727	83,479			
8	,539	4,490	87,969			
9	,497	4,145	92,114			
10	,439	3,658	95,772			
11	,293	2,441	98,213			
12	,214	1,787	100,000			

*Matriz de correlaciones (KMO=.44; prueba de Bartlett: chi-cuadrado 62,30, gl. 66, sig.=,60).

Las comunalidades de las variables tampoco en este caso son excesivamente altas, aunque aceptables. Los términos menos explicados son sexo y guerra (.34 y, 21 respectivamente) y los más altos: evitar (.76), niño (.73), el resto se sitúan entre, 57 y ,67; salvo ciencia con ,51.

Se utilizó el mismo procedimiento y criterio de extracción que en la escala anterior. Hasta el 1 de varianza resultaron cinco factores, con un 68,64 de la varianza. Optamos por la estructura factorial de cuatro factores que dispersaba menos las variables de los componentes (el % de varianza explicada es de 58,85).

En la presentación de la matriz factorial rotada omitimos las cargas factoriales inferiores a .30 para facilitar una comprensión rápida de los componentes.

Los cuatro factores extraídos recogen los trece ítems y sus cargas factoriales oscilan entre .60 y .80, salvo tres variables, sexo y ciencia y guerra, siendo ésta última la que tiene un peso poco significativo.

Tabla 8 (continuación)
Matriz de componentes rotados(a)

	Componente			
	1	2	3	4
niños	,786			
obedecer	,744			
respeto	,679	,377		
evitar		,865		
gente		,814		
deber	,508	,540		
jóvenes			,808	
idea			,797	
persona				,792
ciencia			-,395	,552
sexo				-,471
guerra	-,282			-,306
Varianza ex.	2,025	2,019	1,563	1,456
%	16,873	16,823	13,022	12,132

Método de extracción: Análisis de componentes principales. Rotación Varimax.

El primer conjunto revela un factor de valores tradicionales que al definirse por el mayor peso del término obedecer (.786) seguido de niños (.744), y algo menos el término respeto (.679), se podría ver como elementos de cierta *visión autoritaria*. El segundo núcleo alude a lo que ya

habíamos detectado en la matriz de contexto de términos, la relación entre deber, gente y evitar. Parece evidente que contiene cierto reparo, precaución o distanciamiento. Atendiendo a los ítems en que aparecen, entendemos que es un factor que recoge aspectos de una concepción negativa de la sociedad; podría verse como *cinismo* o bien como prejuicio. Estos dos factores son claves importantes para interpretar el autoritarismo que está implícito en la escala F, al margen de su validación empírica. De los otros dos componentes, quizá el cuarto, aunque es el que menos explica la varianza, es también clave en el concepto del autoritarismo. Alude a la idea de *Irracionalidad*. Lo forman sexo, persona ciencia y guerra, salvo que sexo y guerra están en el polo negativo y en el positivo, persona y ciencia. Es un factor de impulsividad frente a racionalidad. El tercer componente está formado solamente por dos términos, joven e idea, que viendo los ítems en los que aparecen se relaciona con el problema de cambio y, en el fondo, en la terminología de la escala sería un componente del *Conformismo*.

De cara a comprobar si lo expuesto y las claves conceptuales que hemos planteado tienen sentido, analizamos las correlaciones de estos términos con los factores extraídos ya empíricamente (con muestras de sujetos) por Pinillos.

En definitiva, parece que los resultados del análisis de contenido interno de algunos elementos (los seleccionados por su frecuencia) de la escala F, revelan en sí mismos ciertos aspectos de las creencias implícitas de la escala, antes de su validación empírica con muestras de sujetos. Existe cierto grado de congruencia entre nuestro análisis interno y la validación empírica de la escala F, realizada por Pinillos.

Comparación de las escalas y conclusiones

Si comparamos los resultados del análisis de contenido interno realizado en las dos escalas de autoritarismo, hay dos cuestiones claves que debemos destacar.

La primera y, quizá, más relevante es que están manejando dos redes conceptuales distintas de Autoritarismo. Mientras que la RWA se centra en un escenario más político e institucional, la escala F lo hace en el marco de valores y sociedad.

Existen unos elementos internos a las escalas, previos y al margen de su validación empírica con muestras de sujetos, que definen elementos de su concepción del Autoritarismo. Altemeyer se centra en el autoritarismo que coarta y limita las opciones personales, su autoritarismo se sitúa en sociedades democráticas y le preocupa que esas sociedades libres estén amenazadas por elementos o indicadores de autoritarismo (Altemeyer, 1988, 2006). Contrapone fundamentalmente lo personal a lo institucional y presta atención a los elementos institucionales que pueden asociarse a un autoritarismo. Al grupo de Berkeley, le preocupa más

las relaciones sociales y los valores de la sociedad tradicional, que estaba cambiando, así como los elementos racionales e irracionales de la misma. En la medida que sus claves son el prejuicio y distanciamiento social, los elementos irracionales (expresados según nuestra selección de términos frecuentes, en sexo y guerra, que son eros y thanatos) persigue un concepto aparentemente más patológico.

En segundo lugar, los resultados son coherentes con las épocas en que aparecen ambas escalas. La RWA en el momento en que parece que el modelo democrático se ha estabilizado, mientras que la F lo hace en un momento delicado cuando se mantiene aún una confrontación ideológica, aunque sea mediante el conflicto encapsulado de la guerra fría, y con la preocupación de que puedan repetirse acontecimientos demasiados recientes provocados por ideologías extremas.

En tercer lugar, en la matriz de datos con la que hemos obtenido estos resultados, convertimos los ítems de las escalas en sujetos a través de los que analizamos unas variables, los términos más frecuentes. Pues bien, analizando la matriz de correlaciones de dichos términos y obteniendo sus temas centrales con el análisis factorial de componentes principales, descubrimos que antes de su validación empírica y al margen de las respuestas de los sujetos a las escalas, los ítems ya están relacionados entre sí, de forma que implícitamente contienen ideas preconcebidas sobre los elementos básicos que definen el autoritarismo.

Datos curiosos que las diferencian, ya a niveles de características formales, es que mientras la RWA menciona constantemente el término autoridad, la F lo hace, pero con tan poca frecuencia que carece de relevancia en cuanto a frecuencia de términos se refiere. Sus estilos son también diferentes, la RWA más repetitiva tanto en los términos como en las relaciones de términos, pero también con menor grado de diferenciación. La F es menos redundante en los núcleos temáticos que le preocupan, al menos desde la selección de términos establecida en nuestro análisis. Presenta frases sucintas, más sencillas y menos reiterativas, con menos palabras, mientras que el estilo de la RWA es más prolijo (ver tablas 1-2 y 5-6, para la comparación de aspectos formales).

Hemos señalado algunos de los aspectos que diferencian a ambas escalas, cabe preguntarse para cerrar nuestro análisis por los aspectos o ideas preconcebidas que comparten.

Para verlo, establecemos una nueva base datos, esta vez recogiendo de las dos escalas juntas, aquellos términos con frecuencia superior a tres, y como en los casos anteriores, sumando en la frecuencia de cada término los de la misma raíz. El resultado es que comparten 12 de los 18 términos seleccionados: autoridad, deber, gente, jóvenes, libertad, moral, obedecer, orden, padre, persona, sexo, sociedad. Les diferencian, entre otros, ciencia, evitar y respeto que solamente aparecen en la F y agitador, costumbres y país que solo aparecen en la RWA. Es decir, la diferencia importante radica en la frecuencia de los términos

compartidos y las asociaciones entre ellos. La autoridad es la cuestión predominante de la RWA, la obediencia y la juventud las de la escala F. Y en ambas, los términos reiterativos casi por igual son: deber, gente y persona. A la RWA le preocupa más la autoridad y la libertad y a la escala F más la obediencia, la sumisión.

Implicaciones para la medida y análisis de ítems de las escalas

Es indudable que los avances en la ciencia en general, en la matemática y la estadística, junto al desarrollo de nuevas tecnologías para la manipulación y elaboración automatizada de datos obtenidos, ha facilitado que tanto los aspectos métricos como los aspectos personales del investigador, es decir, las creencias previas de todo investigador queden más o menos neutralizadas o, a veces, encubiertas, cuando ha sido necesario hacerlo.

Sin embargo, la construcción de escalas es un viejo tema dentro del campo de la Psicología Social. Ya Thurstone en 1928 alertaba de que la medida de las actitudes debe antes de nada clarificar los términos y suposiciones básicas de las que parte. Además enumeraba una serie de directrices (criterios informales, decía el autor) para la selección de las afirmaciones o preguntas que deberán someterse a depuraciones sucesivas hasta llegar a una selección de las más pertinentes.

La evaluación de cualquier aspecto del ser humano a través de pruebas (sean pruebas objetivas, escalas de actitudes, de rendimiento, de inteligencia o test de personalidad, etc.), implica una serie de pasos previos relacionados con la construcción del propio sistema de evaluación. Las sucesivas decisiones que toma el autor o autores en la elaboración final del sistema o instrumento de medida son tanto o más relevantes que la cuantificación e interpretación de los resultados de los sujetos.

Existen restricciones conceptuales y de escalas de medida a la hora de detectar, cuantificar e interpretar las actitudes. Inicialmente se orientaron a la predicción de conductas futuras, enfrentándose de entrada a los procedimientos adecuados para su construcción y posterior validación (fiabilidad y los distintos tipos de validez) (Summers, 1976). Aunque se ha dado prioridad a la última cuestión, el primer paso, la elaboración de las preguntas determina en gran medida su alcance, significado y, sobre todo, la generalización de los resultados obtenidos (no en el sentido tradicional de generación de la muestra a la población) sino la generalización del concepto de lo medido.

Robinson y otros (1993) en su libro *Measures of Political attitudes enumeran* algunas cuestiones metodológicas relativas a la calidad de las escalas de actitudes. Señalan tres pasos iniciales que son de especial relevancia en su construcción de las escalas:

En primer lugar, la selección de las preguntas pertinentes y adecuadas al campo que se quiere estudiar y para el que se pretende construir una escala. Realizada la selección, otra cuestión es encontrar la forma más adecuada de expresar dichas preguntas y, por último, el análisis de ítems. Solamente después nos introducimos en otras dos cuestiones: los problemas de la aquiescencia y la deseabilidad que pueden introducir errores de medida según el comportamiento de las nuestras con las que se validó la escala. En cualquier caso, es inevitable la depuración de la selección previa y nueva contrastación de dichos efectos artificiales de medida. En tercer lugar y, resueltos los dos anteriores, ya es el momento de plantearse los problemas de fiabilidad y validez de la escala construida.

El objetivo de este trabajo está precisamente en el primer conjunto de problemas señalados por Robinson (1993) y otros tantos autores; la selección, formulación y análisis de ítems.

Cuando un investigador pretende medir un aspecto psicológico, o un acontecimiento del momento con efectos psicológicos sobre una población, o las actitudes ante una determinada política social, se enfrenta a una dificultad inicial: ¿Dónde buscar? Existen pocos fenómenos sobre los que no haya definiciones, elaboraciones conceptuales, listados de características relacionadas e incluso pruebas de medida y resultados ya reelaborados en teorías. En el ámbito de las ciencias sociales esta cuestión es además compleja porque es raro encontrar un acuerdo de definición, enfoques o dimensiones detectadas del problema. El listado de literatura es tan amplio que el investigador se ve obligado a realizar una selección y delimitación de tanto conocimiento acumulado. Nos enfrentamos a la *primera inferencia*, determinada en gran medida, por esas creencias básicas de las que hablamos al principio y que caracteriza a todo investigador. En *la elección de ítems* se utilizan distintos procedimientos para seleccionar aquellos que cubran mejor el fenómeno que se pretende medir, pero al final, siempre existe un *margen discrecional* de los propios investigadores, que raramente desvelan o explícitamente describen.

Otra cuestión es *la expresión verbal de los ítems* elegidos. Anastasi (1977), en su libro *Test psicológicos*, dedica un capítulo al análisis de elementos. Aunque se centra en pruebas de rendimiento señala algunos procedimientos y guías publicadas para mejorar su construcción. Pueden encontrarse más directrices en Thurstone y Chave (1929), Cronbach y Gleser (1965), Converse y Presser (1986), Sudman y Bradburn (1982) por no mencionar las sucesivas directrices publicadas por la APA para la selección y redacción eficaz de los elementos de una escala.

Los métodos para solucionar el problema del *Análisis de los ítems* de una futura escala son muy diversos. Desde el análisis factorial, hasta los análisis multidimensionales son procedimientos estadísticos de gran uti-

lidad para pulir los ítems o elementos definitivos que forman la escala de medida que se está construyendo.

En estos dos últimos aspectos señalados, los procedimientos utilizados pasan por pruebas piloto con muestras de sujetos que sirven para detectar tanto las expresiones incorrectas de los ítems, su redundancia o ausencia de relación con el resto de los ítems seleccionados para, después de haber pulido el banco de preguntas, realizar las pruebas pertinentes de validez y fiabilidad con muestras de sujetos adecuadas.

El trabajo presentado aquí pretende poner a prueba un procedimiento estandarizado de análisis interno de las escalas construidas y, por tanto, útil para la depuración de ítems antes de llegar a comprobar su fiabilidad y validez empíricamente, con muestras de sujetos reales. No solo debido a que indudablemente, la utilización de pruebas y muestras piloto añaden nuevos factores que sesgan de forma externa el contenido de los ítems en su depuración, sino fundamentalmente porque los autores que deciden construir una escala de actitudes siempre se ven obligados a utilizar algún criterio selectivo de lo que quieren preguntar. Sean o no conscientes, dicho criterio lleva consigo ideas preconcebidas que determinan los límites conceptuales de la escala construida. Es decir, lo realmente importante para la validez posterior con muestras de sujetos, es precisamente la validación del concepto previo que los autores tienen sobre el tema investigado.

Pretendemos justificar y presentar un procedimiento que ayude a mejorar y sobre todo a encontrar un sistema lo más automático posible, de forma que cualquier investigador pueda realizar el análisis de los ítems de una escala y llegar a los mismos resultados que otros investigadores. Nuestro análisis interno recoge los problemas de la selección de ítems y su expresión verbal, tal como señalan los autores que se han enfrentando desde hace tiempo a la construcción de escalas.

Lo hemos hecho utilizando, a modo de ejemplo, dos escalas sobre una misma problemática. El análisis de las escalas RWA y F pone de manifiesto que ambas contienen en sí mismas los gérmenes de lo que van a detectar. Proponemos un procedimiento de análisis interno de las escalas para determinar su alcance y generalización conceptual, más allá y antes de su validación empírica con muestras de sujetos.

Referencias

- Adorno, T.-Frenkel-Brunswick, E.-Levinson, D.J.-Sanford, R.N.(1950): *The authoritarian personality*. New York: Harper and Row
- Altemeyer, B.(1981): *Right-Wing Authoritarianism*. Winnipeg: Univ. Manitoba Press.
- Altemeyer, B.(1988): *Enemies of Freedom. Understanding right-wing authoritarianism*. San Francisco: Jossey-Bass

- Altemeyer, B. (2006): *The Authoritarians*. Canada. Manitoba. (formato en "pdf", Copyright ©, 2006 by Bob Altemeyer)
- Anastasi, A. (1977): *Test Psicológicos*. Madrid: Aguilar, 3ª ed. (orig. 1968).
- Converse, J.-S. (1986): *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage
- Cronbach, L.J.-Gleser, G.C. (1965): *Psychological tests and personal decisions*. Urbana III: University of Illinois Press, 2ª ed.
- Garcés, J. (1985): Sistema de valores en política autonómica: perfil ideológico de discursos políticos, 1984-85, Valencia, *Doctorado*. Dir. J. Seoane
- García de las Bayonas, A. (1983): Análisis del Contenido de un Medio de Comunicación Masiva: El País, 1982-83, Valencia, *Licenciatura*, Dir. J. Seoane
- Garzón, A.-Herrera, M. (1988): Psicología y Política: la racionalización de la Guerra. Valencia: *Boletín de Psicología*, nº 21, 51-84.
- Garzón, A. (1992): Marcos Sociales de la Memoria: un enfoque ecológico. En D. Barcia (Ed): *Trastornos de la Memoria*. Barcelona: Editorial MCR, págs. 79-101.
- Garzón, A.-Seoane, J. (1982): La memoria desde el Procesamiento de Información. En I. Delclaux, J. Seoane (Eds.): *Psicología Cognitiva y Procesamiento de Información*. Madrid: Pirámide, 117-140.
- Garzón, A.-Seoane, J. (1996): *Técnicas y Prácticas Instrumentales en Psicología*. Valencia: Colección Psicología Teorética, Promolibro.
- Herrera, M. (1986): Violencia Política: Hacia un modelo psicológico en la teoría de la Guerra, *Doctorado*. Dir. A. Garzón.
- Krippendorff, K. (196): *An Examination of Content Analysis: A Proposal for a Framework and an Information Calculus for Message Analytic Situations*, Ph.D. Dissertation, Urbana: University of Illinois
- Krippendorff, K. (1986): *Information Theory: Structural Models for Qualitative Data*, Beverly Hills, CA: Sage Publications
- Krippendorff, K. (1980): *Content Analysis. An introduction to its methodology*. London: Sage.
- Lindsay, P.H.-Norman, D.A. (1975). *Procesamiento de Información Humana*. Madrid: Tecnos, Traducción de J. Seoane-C. García-Tervijano, (orig. 1972)
- Moreno, J.J. (1985): Estudio diferido de Psicología Política: análisis psicosocial de textos políticos de D. Juan Lerma Blasco, 1984-85, Valencia, *Licenciatura*. Dir. J. Seoane
- Moreno, J.J. (1987): Psicobiografía Política de Enrique Tierno Galvan, 1986-87, Valencia, *Doctorado*. Dir. J. Seoane
- Pinillos, J.L. (1963): La Escala "F" en una muestra española: estado comparativo adaptación original de F. *Psicología General y Aplicada*, vol, XVIII, 70, 1155-74.
- Robinson, J.P.-Shaver, P.H.-Wrightman, L.S. (ed.) (1993): *Measures of Political Attitudes*. Londres: Academic Press.
- Rokeach, M. (1960): *The open and closed mind*. Nueva York: Basic Books, Inc.
- Sanfeliu, A. (1986), Sistema de valores en las Elecciones Generales de 1977: espacios radiofónicos de inserción obligatoria difundidos por las emisoras valencianas, 1985-86, Valencia, *Licenciatura*. Dir. J. Seoane
- Sanfeliu, A. (1988): La Radio como medio de comunicación social, 1987-88, Valencia, *Doctorado*. Dir. J. Seoane.
- Seoane, J. (1979): Inteligencia Artificial y Procesamiento de la Información; *Boletín Informativo* 85, 3-21. *Fundación Juan March*
- Seoane, J. (1988): *Psicometría*. Madrid: UNED 4ª ed., (1976 1ª ed.)
- Seoane, J.-Garzón, A. (1992): Creencias Sociales Contemporáneas, Autoritarismo y Humanismo. *Psicología Política*, 5, 27-52

- Strauss, A.(1987): *Qualitative Analysis for Social Scientists*. Cambridge: Cambridge University Press.
- Sudman,S.-Bradburn,N.(1982): *Asking question*. San Francisco: Jossey Bass.
- Summers, G.F.(1976): *Medida de las Actitudes*. Méjico: Trillas (orig., 1970).
- Thurstone,L.L (1928): Attitudes Can Be Measured. *Am. J. of Sociology*, 33, 529-54.
- Thurstone,L.L.-Chave,E.J.(1929): *The measurement of attitude*. Chicago: University of Chicago.