

Content Validity and Metric Properties of a Pool of Items Developed to Assess Humor Appreciation

Hugo Carretero-Dios, Cristino Pérez, and Gualberto Buela-Casal
Universidad de Granada (Spain)

This research deals with assessing humor appreciation and highlights some of the strategies that can be used in two necessary stages in the construction of a test: the content validity study and the item analysis. First, we analyzed the content validity of a battery of 200 items developed to assess humor appreciation. Second, we analyzed the metric properties of the selected items by means of two studies. The first study was a pre-pilot analysis of the items in a sample of 212 participants, and the second study was a new item analysis in a sample of 344 Spanish people aged between 18 and 71 years. To determine content validity, we calculated interjudge agreement on item-facet theoretical match. Each item was assessed by seven judges, and the selection criterion used was a minimum agreement of 70%. This procedure led to eliminating 27 items.

Subsequent item analyses led to a preliminary proposal for a 40-item scale (*Escala de Apreciación del Humor*, EAHU [Humor Appreciation Scale]) with appropriate descriptive statistics as well as discrimination and homogeneity values. The internal empirical structure of the scale matched the operative definition of humor appreciation, and the Cronbach's alpha of the EAHU scores ranged from .72 to .89.

Keywords: humor appreciation, content validity, item analysis, test construction, EAHU.

El presente estudio aborda la problemática de la valuación de la apreciación del humor, a la vez que se pormenorizan algunas de las estrategias que pueden usarse dentro de dos de las fases de necesaria ejecución dentro de todo proceso de construcción de un test: el estudio de la validez de contenido y en el análisis de ítems. En primer lugar, se analizó la validez de contenido de una batería de 200 ítems desarrollada para evaluar la apreciación del humor. En segundo, y a través de dos estudios independientes, se analizaron las propiedades métricas de los ítems seleccionados. El primer estudio se trató de un análisis pre-piloto de los ítems a través de una muestra de 212 participantes, y el segundo fue el análisis de ítems propiamente dicho a partir de una muestra de 344 participantes españoles con edades comprendidas entre los 18 y los 71 años. Para determinar la validez de contenido se calculó el acuerdo inter-jueces sobre el grado de pertenencia teórica ítem-faceta. Cada ítem fue evaluado por siete jueces y el criterio de selección se basó en un nivel de acuerdo mínimo del 70%. Este procedimiento condujo a la eliminación de 27 ítems. Los consiguientes análisis de ítems derivaron en una propuesta de escala preliminar de 40 ítems (*Escala de Apreciación del Humor*, EAHU) caracterizados por presentar unos adecuados estadísticos descriptivos a la vez que unos valores de discriminación y homogeneidad apropiados. De igual forma, la estructura interna de la escala se correspondió con la definición operativa de la apreciación del humor, presentando los distintos factores valores alfa de Cronbach que oscilaron entre 0,72 y 0,89.

Palabras clave: apreciación del humor, validez de contenido, análisis de ítems, construcción de tests, EAHU.

In research about sense of humor, the assessment of humor appreciation has been considered important for a long time. In fact, the first instruments developed to assess sense of humor focused on humor appreciation (Cattell & Luborsky, 1947; Cattell & Tollefson, 1966; Eysenck, 1943) and measured the degree of funniness shown by subjects when they were shown comic cartoons. Results available to date have shown that this dimension has a two-fold importance; from a conceptual point of view, it grants theoretical support to general models of personality (Galloway & Chirico, 2008; Ruch & Hehl, 1998); it is also considered to be important in applied research, as can be seen in very diverse areas of work (Bing, 2007; Bozikas, et al., 2007; Eyssel & Bohner, 2007; Ford, Boxer, Armstrong & Edel, 2008; Samson & Huber, 2007).

In spite of the above, the attention given to humor appreciation has not led to good quality assessment tests. Such tests have numerous shortcomings, the most outstanding of which are the following: 1) the instruments are usually developed for a particular study and the construction strategy is not explained; 2) the basic psychometric properties of items are unknown or not presented; 3) the scales consist of a group of humor stimuli – jokes or cartoons – selected by the authors according to certain theoretical humor categories, but not subjected to empirical analysis to corroborate the internal structure of the group of jokes; 4) validity studies are not carried out to assess the relevance of the categories proposed (for greater details, see Carretero-Dios, Pérez & Buela-Casal, 2006).

The 3 WD Humor Test, 3 WD (Ruch, 1992) is the exception to the previous statements on assessment tests. This scale is supported by numerous research findings and has a clearly defined conceptual framework (Ruch & Hehl, 1998). Based on the approach of Ruch's group, a dimensional proposal or semantic definition (Lord & Novick, 1968) of the humor appreciation construct (Carretero-Dios et al., 2006) was recently put forward to correct some of the defects mentioned. This proposal was subjected to an evaluation process by experts. Its results, along with the theoretical and empirical information considered (Carretero-Dios et al., 2006), led to the acceptance of the proposal as a basis for the creation of a new assessment scale: *Escala de Apreciación del Humor* (Humor Appreciation Scale), EAHU.

Basically, and following the contributions made by Ruch and his group (Ruch, 1992; Ruch & Hehl, 1998), we understand humor appreciation as the degree of *funniness* and *aversiveness* experienced as a response to a given humor stimulus. In the specialized literature, only responses with a positive valence (funniness) are considered. However, a factor analysis of affective responses to humor yielded the two orthogonal dimensions of positive and negative responses (Ruch, 1992). Aversiveness is conceptualized as the degree to which a joke is considered inappropriate, annoying, offensive, etc., depending of its characteristics.

In the dimensional proposal on humor appreciation, a distinction is made between the content of the humorous material (no specific content, sexual humor, black humor and disparagement humor) and the cognitive processes that are triggered when such material is perceived (incongruity-resolution and nonsense; see Ruch & Hehl, 1998). In incongruity-resolution humor (INC-RES), a two-stage process can be observed: the perception of an incongruity and its resolution. In this type of humor, an incongruity is discovered and then resolved using the information available elsewhere in the joke or cartoon. Although nonsense humor (NON) also has an incongruous punch line, “the punch line may 1) provide no resolution at all; 2) provide a partial resolution (leaving an essential part of the incongruity unresolved), 3) or actually create new incongruities” (McGhee, Ruch & Hehl, 1990, p. 124). The fact of cross-matching the contents – not all the possible ones, only those selected from a theoretical point of view – with the cognitive processes is supposed to lead to the operative components of humor appreciation (see Appendix 1). However, it would be appropriate, as far as possible, to empirically separate the content of the humorous material from the cognitive processes (see Carretero-Dios et al., 2006 for the theoretical and empirical justification). Hence, the operative proposal on humor appreciation would be formed by five facets: 1) incongruity-resolution humor (INC-RES); 2) nonsense humor (NONS); 3) sexual humor (SEX); 4) black humor (BLACK); and 5) disparagement humor (DIS), which was defined as sexual disparagement humor for theoretical purposes.

The objective of this study was to reach consensus on the scientific definition of humor appreciation, obtain the necessary theoretical validity evidence for the definition through the judgment of experts (Carretero-Dios, et al., 2006) and finally follow the logical process of the creation of an evaluation instrument (AERA, APA & NCME, 1999). We planned to study the content validity of a battery of 200 items developed to assess humor appreciation and study the metric properties of the items in a second stage. This stage involved two studies: the first study was a pilot analysis of the items (Henrysson, 1971) selected from the content validity assessment, and the second study involved using a larger sample and repeating the analyses on a subset of selected items to develop the final scale.

This research is aimed at providing information that can be used as an applied and methodological contribution. Thus, we do not only wish to provide some data – content validity and item analysis – on a new evaluation instrument, but also to point out some of the possible strategies to adopt when dealing with two of the necessary stages in the construction of a psychological assessment test: the content validity study and the analysis of the metric properties of the items. For this reason, we will introduce the central aspects of both stages and their purpose in the construction of a new measuring instrument before presenting the results obtained for both stages.

Item construction and content validity study

Content validity is defined as the extent to which items adequately samples the domain of interest when attempting to measure construct (Polit, Beck & Owen, 2007), that is, whether the domain of content for the construct is adequately represented by the items (Haynes, Richard & Kubany, 1995). The content validity study of the items developed for a given assessment scale represents a critical early step in enhancing the construct validity of an instrument (Mastaglia, Toye & Kristjanson, 2003).

The process aimed at studying the content validity of an instrument has been divided into two major stages (Sireci, 1998). The first stage involves clearly defining the construct of interest by specifying its domains of content operationally. Such domains must be derived from a broad review of specialized literature and agreed on the basis of an assessment by a group of experts on the subject (see this stage in Carretero-Dios et al., 2006). The second stage involves developing the items that are valid operational indicators of the domains of content agreed for the construct, and determining to what extent each item is relevant for the component of the construct it was created for by means of an assessment by judges.

Item construction stage

A table of item specifications was developed on the basis of the dimensional proposal on humor appreciation (Spaan, 2006). This table serves as guidance for the next steps to take in the process of constructing a new assessment tool (Gordon, 2004; Haladyma, 2004), and especially for the operational concretion of the construct through the development of an item pool. Using the table of specifications of the items as a guide, one of the authors was in charge of selecting a broad item pool (from the Internet, books, newspapers and other sources). In all the cases, the items were written jokes and cartoons. After discussion with the rest of the authors, an initial pool of 200 items – 40 for each type of humor – was proposed. The same number of items was chosen for man disparagement and woman disparagement humor. We used different criteria to select the items: 1) the three authors had to agree on which humor appreciation component each joke or cartoon corresponded to; 2) the jokes or cartoons should not refer to any specific social or historic event. This was done to ensure that the understanding of the selected material did not depend on specific knowledge about the facts, and therefore that the information provided by the jokes or cartoons themselves was sufficient; 3) the language of the jokes or cartoons should make it possible to assess a sample with a broad age range; it should therefore not be based on youth jargon, expressions of “other” generations, or the like (see an example of one item per dimension in Appendix 2).

The judgment-quantification stage

The content validity study was aimed at carrying out an analysis without formal statistical criteria for interpretation (Carretero-Dios & Pérez, 2005; Haynes, et al., 1995; Nunnally & Bernstein, 1995). Instead, the intention was to use procedures that make it possible to estimate in “degrees” (Henson & Douglas, 2005) whether the items created are conceptually appropriate for the target facet (Beck & Gable, 2001; Mastaglia, et al., 2003; Sireci, 1998). Following the work of Crocker, Llabre and Miller (1988), we decided to work with 7 judges per item.

Method

Participants

We followed the considerations of Grant and Davis (1997) to select the experts. Given the characteristics of the task, the criteria we used to select our judges were that they should be psychologists with experience in the construction/adaptation of evaluation instruments. Given the high number of items, we produced five booklets of different elements (see the Instruments section), and therefore required 35 judges. The judges were mainly lecturers and PhD fellows in the School of Psychology of the University of Granada, Spain.

Instruments

- Table with the semantic definition of humor appreciation (see Appendix 1).
- Assessment booklet.

Given the high number of items, we decided to distribute the items randomly into a total of five booklets (forty items per booklet, with an equal number of items per facet). Each booklet began with a set of detailed instructions about the task the judges were required to perform. The order of presentation of the elements in each of the booklets was decided randomly.

Procedure

First, each judge was given a cover letter asking him/her to cooperate and explaining the purpose of the task and the reason for being selected. Once the judges accepted to cooperate, they were randomly assigned one of the five booklets and a table with the semantic definition of humor appreciation was given to each judge. A matching task was used to obtain the judges answers. They were asked to say which facet of the construct they believed each item corresponded to. They were given a space to make qualitative considerations about the clarity and understanding of the task and any relevant issues (see the instructions given to the judges in Appendix 3).

To analyze the answers, was used a derivation of the Content Validity Index, CVI (Wynd, Schmidt & Schaefer, 2003). For the purposes of this study, the CVI was the proportion of judges that matched the item to its intended facet. We decided to rule out any items that had not been classified by at least 5 out of the 7 judges into the intended category of humor appreciation (CVI > .70). As regards the humor contents, we only considered whether the items were included or not in the content of interest, without considering the cognitive process that would also serve to define them (see Carretero-Dios et al., 2006 for a theoretical justification). The qualitative information provided by the judges was to be jointly discussed by the authors until consensus was reached on possible improvements.

Results

The 35 judges consulted completed the content validity task without pointing out any difficulties in understanding the instructions or mentioning errors or formal issues that should be corrected in the items assessed. The data obtained from the judges' assessment of item-facet match led to eliminating 27 items (CVI < .70), which reduced the initial pool to 173 elements. For each of the humor appreciation categories included in the semantic definition of the construct, the number of items discarded was the following: sexual humor (2 items); disparagement humor (3 items); black humor (6 items); incongruity-resolution humor (8 items); and nonsense humor (8 items).

Out of the 173 items selected, 22 obtained a CVI of .71, 125 items had a CVI of .86, and for 26 items there was total agreement between judges, and thus a CVI of 1.

Item analysis

Item analysis is aimed at exploring the individual metric quality of each of the items constructed to be part of a scale (Muñiz, Hidalgo, García-Cueto, Martínez & Moreno, 2005; Osterlind, 1989). The metric quality of items is known to have a direct influence in the final quality of the test as an instrument that provides valid and reliable scores. Thus, the final objective of studies aimed at statistically analyzing items is to select the most appropriate elements based on the most relevant metric parameters according to the conceptual approach to the construct assessed.

According to the research carried out by Henrysson (1971), which is included and quoted in several specialized monographs (Carretero-Dios & Pérez, 2005; Clark & Watson, 2003; Martínez-Arias, 1995), there are two important stages in item analysis that can be considered pilot stages: the pre-pilot stage, with all the items constructed, and the analysis stage itself, which uses items that were not discarded

in the pre-pilot stage or a selection of items among those considered to be suitable.

The pre-pilot stage is considered to be the construction stage of a test, when the battery of items available is used for the first time with a sample ranging between 50 and 100 participants. The purpose is to perform the first analysis of the items, looking at statistical parameters (quantitative analysis) and formal aspects.

In our work, we followed the objectives of this pre-pilot study. As regards formal aspects, the objectives were: 1) to determine the mean time interval it took the participants to answer the items; 2) to check whether the instructions were correctly understood, and 3) to verify whether the test was answered correctly according to the design of the instrument. For the metric analysis of the items, the aims were: 1) to analyze the degree of adherence to the items, the measuring range the items were sensitive to and the behavior of the response options; 2) to determine the capacity of each item to discriminate between subjects with high and low scores in the target facet, and 3) to analyze the homogeneity of the items in terms of their ability to differentiate themselves from items from facets other than their own. We planned to use the analysis of the criteria mentioned to select a smaller battery of items with the most appropriate metric and formal properties possible.

The item analysis stage is the logical continuation of the pre-pilot stage (Carretero-Dios & Pérez, 2007). It involves working with a larger sample to confirm the results obtained in the pre-pilot stage. The recommendation is to use a sample ranging between 5 and 10 participants per item (Carretero-Dios & Pérez, 2005) and include calculation procedures that are appropriate for the sample size. Thus, this would be the first stage at which factor analysis would be recommended in the construction of a new scale, as a strategy focused on item analysis. The purpose would be to explore the empirical "separation" between the items of the scale.

Next, we shall present two independent studies – the pre-pilot item analysis and the pilot item analysis – of the 173 items available for the development of the EAHU. The objectives were the ones specified for both stages in the previous paragraphs.

Study 1: *Pre-pilot item analysis*

Method

Participants

We used a sample of 212 participants, 108 males (mean age 25.18, standard deviation 4.09) and 104 females (mean age 23.93, standard deviation 3.13). The sampling procedure was non-probabilistic (see the Procedure section). Out of the total sample, 154 participants were university students, and 57 were not university students.

Instruments

We used 173 items. The items were rated on 2 unipolar 5-point scales for “funniness” (from 0 = not at all funny, to 4 = very funny) and “aversiveness” (from 0 = not at all aversive to 4 = very aversive). The number of items per dimension was the following: SEX_(sexual humor) = 38 items; BLACK_(black humor) = 37 items; DIS_(sexual disparagement humor) = 37 items; INC-RES_(incongruity-resolution humor) = 32 items; NONS_(nonsense humor) = 32 items.

Given the high number of items, we constructed four booklets of elements, randomly allocating the same number of items to each component. The order of appearance of the items in each booklet was also decided randomly.

Procedure

We decided to operate in different libraries of the University of Granada. Data collection was done as follows: at the library, a person who had received prior instruction – always the same individual – asked the people present to cooperate in “a study on sense of humor that was being carried out at the University of Granada and involved reading a series of comic cartoons.” People who decided to cooperate were randomly allocated one of the booklets of elements by means of a table of random numbers (Booklet 1 $n = 49$; Booklet 2 $n = 54$; Booklet 3 $n = 61$; Booklet 4 $n = 48$).

Statistical analyses

For the metric analysis of the items, calculations were made for each facet of humor appreciation, and the results were grouped depending on the funniness or aversiveness responses. The following calculations were made:

Basic descriptive statistics. For each of the items, the minimum, maximum, response range, mean and standard deviation were calculated.

Discrimination index. Two procedures were used, the first of which was the corrected item-total correlation. The second one was to select the subjects with a value equal to or lower than the 27th percentile and equal to or higher than the 73rd percentile in their funniness and aversiveness scores for each item, and then calculate the differences in the mean scores between both groups. Two observations should be noted about the second procedure. First, the cut points selected for both groups (27th and 73rd percentile) were the values recommended to study item discrimination when the purpose is to work with “high” and “low” groups in any dimension from a statistical point of view (Muñiz, et al., 2005). The second observation explains the use of this second procedure as a discrimination index. This procedure is used to confirm the results obtained with the corrected item-total correlation procedure. Given the sample sizes in the pre-pilot stages of item analysis, it is

recommended to use at least a second calculation procedure that is more appropriate for such sample sizes and can support the data. In these cases, the recommendation is to compare the extreme groups in the dimensions analyzed (Osterlind, 1989).

With these indicators, we eliminated a first set of elements, namely those with the following characteristics: 1) a response range below 3; 2) a standard deviation below 1 (this latter criterion was not applied for aversiveness scores to components with no specific content, given that we did not expect high mean aversiveness regarding this kind of humor, nor dispersion in the responses); 3) a corrected item-total correlation below $r = .30$; and 4) a positive or zero value in the difference between the scores of the participants situated at the extremes.

After discarding the relevant elements under the criteria specified, we analyzed the homogeneity of the items. For this purpose, we calculated the correlation value of each of the items with the total value of its own dimension and the total value of the remaining dimensions. This was aimed at selecting items that feature higher correlation values with their own dimension than with any of the remaining ones (Jackson, 1970). The criterion established was to rule out elements whose correlation with the total value of their dimension did not exceed their correlation with any of the other dimensions by at least two tenths (Jackson, 1970).

Results

The pre-pilot study of the metric properties of the items was determined by the use of the four different samples of participants, who responded to four different booklets of items. This led us to explore first whether there were statistically significant differences between the different groups of participants in the funniness and aversiveness scores. It is worth noting that it would be desirable not to find any differences between the groups, which would make it possible to compare the statistical results between the groups with greater certainty. Five analyses of variance were carried out for the funniness scores, and another five were performed for the aversiveness scores. The fixed factors considered were the different groups of participants, and the dependent variable was the total score in each of the five dimensions of humor appreciation. No statistically significant differences were found between the four groups of participants for the dimensions studied.

Out of the 173 initial items, 87 were finally selected. There were 19, 19, 17, 16 and 16 items to measure sexual, sexual disparagement, black humor, incongruity-resolution and nonsense humor respectively. It must be noted that the analysis that led to discarding the greater number of items was the one used to calculate homogeneity, that is, the correlation of each of the elements with the total value of its own dimension (corrected correlation) and with the

total value of the remaining dimensions. In general, all the elements correlated more closely with their own dimension than with the rest, although they related positively with dimensions that were not their own. In the next study we intend to work with a larger sample common to all items and explore the relations between the dimensions of humor appreciation in greater detail.

Study 2: Item analysis of a first proposal of the Escala de Apreciación del Humor (Humor Appreciation Scale), EAHU

Method

Participants

The sample included 344 Spanish people, 160 males (from 18 to 71 years; $M = 36.76$, $SD = 12.50$), and 184 females (from 18 to 66 years; $M = 31.15$, $SD = 11.33$). Out of the total sample, 108 were university students from the University of Granada, and the others were adults from the region.

Instruments

A first proposal of the EAHU was used with the items selected from the pre-pilot item analysis. The scale was formed by 40 items (EAHU₄₀: 8 for each category of humor appreciation). The items of each category were selected randomly from the total of those available for each one. Randomness only had one restriction for the sexual disparagement component (equal number of items for man disparagement and for woman disparagement humor). The order of appearance of the items on the scale was also decided randomly. The items were rated on 2 unipolar 5-point scales for “funniness” (from 0 = not at all funny, to 4 = very funny) and “aversiveness” (from 0 = not at all aversive to 4 = very aversive).

Procedure

In some libraries of the University of Granada (Spain), the EAHU₄₀ was handed out to those people who wished to collaborate. The scale was preceded by standard instructions and a consent form. Along with this sample-collecting procedure, we used quota sampling based on sex (similar number of men and women), and age (over thirty years old).

Statistical analysis

We made the same calculations as in the pre-pilot study (see the section on Statistical analysis in Study 1), except the comparisons between the extreme groups, given the size

of the sample in this study. Besides, we applied exploratory factor analysis to the participants' responses using principal component analysis (PCA) with Varimax rotation.

Results

First, we explain how the items of the scale will be designated in the presentation of the results. The label used for the dimension appears first (SEX, BLACK, DIS, etc.). Next to the label there is a number – from 1 to 40 – that corresponds to the place of the item in the scale.

Descriptive statistics and discrimination indexes

Table 1 shows the descriptive statistics and discrimination index for the funniness and aversiveness responses.

Table 1 shows that the descriptive statistics and corrected item-total correlations were confirmed for this sample. No items showed a corrected item-total correlation $< .30$ in any of the two dimensions. Items with options that had not been chosen or with a response range < 3 did not appear.

Internal structure of EAHU₄₀

A principal component analysis with Varimax rotation was performed. The KMO value was 0.90, and Bartlett's test showed statistical significance ($Chi-square = 4204.40$, $df = 496$, $p < .01$), indicating that the samples met the criteria for the factor analysis. Seven factors exceeded unity (Eigenvalues were 3.75, 3.44, 3.22, 2.27, 2.09, 2.03 and 1.89). Nevertheless, Factor VII was mainly defined by items with loadings $< .30$. The items in this seventh factor with loadings $> .30$ also had higher loadings in some of the other six factors. The Scree test suggested retaining six factors, which explained 52.51% of the variance. The factors retained after Varimax rotation are shown in Table 2.

The factors were identified as black humor (factor I), sexual humor (factor II), nonsense humor (factor III), man disparagement humor (factor IV), woman disparagement humor (factor V) and incongruity-resolution humor (factor VI). As can be observed in Table 2, the factor analysis separated the items of sex disparagement humor into two different components, one for man disparagement humor (factor IV), and one for woman disparagement humor (factor V).

Scale reliabilities and intercorrelations

Intercorrelations among the EAHU₄₀ scales and Cronbach's alphas were computed. The results are shown in Table 3. Based on the results of the factor analysis, we decided to distinguish between man disparagement humor (M-DIS) and woman disparagement humor (W-DIS) depending on whether men or women were disparaged in the factor male disparagement humor.

Table 1
Descriptive statistics and corrected item-total correlation for the EAHU items

Items	Funniness			Aversiveness		
	M	SD	CITC	M	SD	CITC
Incongruity-resolution humor						
INC-RES1	2.12	1.29	.30	.95	1.01	.43
INC-RES2	2.21	1.22	.31	.99	1.12	.52
INC-RES17	2.67	1.09	.43	.34	.76	.40
INC-RES18	2.54	1.18	.43	.21	.60	.52
INC-RES25	2.57	1.19	.44	.24	.65	.57
INC-RES35	1.98	1.16	.48	.38	.80	.51
INC-RES36	1.68	1.21	.39	.68	.98	.48
INC-RES40	2.78	1.06	.44	.67	.99	.51
Nonsense humor						
NON12	1.04	1.23	.50	.31	.84	.72
NON13	1.01	1.23	.61	.22	.72	.79
NON14	1.51	1.20	.52	.23	.71	.73
NON19	1.53	1.19	.52	.17	.55	.62
NON20	.89	1.13	.56	.27	.78	.61
NON27	1.16	1.21	.53	.28	.69	.67
NON28	.99	1.19	.44	.42	.95	.57
NON29	1.34	1.16	.33	.39	.91	.61
Sexual humor						
SEX3	2.01	1.32	.47	1.38	1.39	.56
SEX4	1.78	1.31	.39	1.09	1.30	.59
SEX10	1.73	1.29	.64	1.21	1.41	.63
SEX11	2.31	1.25	.44	.92	1.14	.68
SEX21	1.81	1.31	.56	1.11	1.39	.57
SEX22	1.69	1.25	.60	1.01	1.20	.65
SEX30	1.92	1.28	.60	.92	1.18	.69
SEX33	2.19	1.18	.50	.95	1.28	.68
Black humor						
BLACK5	1.27	1.32	.57	2.87	1.41	.65
BLACK6	.87	1.23	.66	3.14	1.31	.58
BLACK8	.98	1.19	.58	2.77	1.44	.58
BLACK15	1.63	1.30	.52	1.42	1.47	.44
BLACK16	2.01	1.38	.46	1.90	1.55	.52
BLACK23	1.24	1.27	.64	2.35	1.56	.62
BLACK32	.99	1.26	.56	2.76	1.45	.63
BLACK34	2.02	1.33	.39	1.36	1.42	.45
Disparagement humor						
DIS7	2.22	1.26	.31	1.01	1.24	.40
DIS9	2.38	1.30	.44	.85	1.25	.44
DIS24	.95	1.17	.48	2.05	1.60	.60
DIS26	1.02	1.22	.43	2.26	1.61	.60
DIS31	1.30	1.31	.46	1.52	1.58	.63
DIS37	1.74	1.21	.49	.94	1.28	.54
DIS38	1.90	1.34	.57	1.40	1.51	.65
DIS39	1.61	1.38	.46	1.46	1.49	.56

Note: N = 344; INC-RES = incongruity-resolution; NON = nonsense; SEX = sexual; DIS = disparagement; CITC = corrected item-total correlation.

Table 2
 Loadings of the EAHU₄₀ items on the six rotated factors

Items	F I	F II	F III	F IV	F V	F VI	<i>h</i> ²
BLACK6	.79						.69
BLACK8	.78						.61
BLACK23	.73						.60
BLACK5	.72						.55
BLACK32	.66						.54
BLACK15	.52		.36				.53
BLACK16	.49						.51
BLACK34	.34						.48
SEX10		.74					.63
SEX30		.72					.58
SEX21		.71			.31		.55
SEX22		.65					.51
SEX33		.58					.50
SEX3		.53					.48
SEX11		.47					.45
SEX4		.33					.39
NON13			.79				.69
NON20			.73				.60
NON27			.69				.57
NON12			.62			.30	.55
NON19			.61				.54
NON14			.59				.51
NON28			.45			.36	.44
NON29			.42				.41
DIS7				.71			.55
DIS9				.71			.55
DIS39		.33		.57			.50
DIS31	.31			.46			.45
DIS38				.45	.66		.54
DIS26	.44				.65		.56
DIS24	.44				.63	.36	.55
DIS37					.60		.53
INC-RES18						.76	.65
INC-RES25						.72	.58
INC-RES17						.68	.56
INC-RES40		.32		.31		.61	.55
INC-RES36						.55	.49
INC-RES25						.52	.48
INC-RES1						.34	.46
INC-RES2						.32	.38
Eigenvalues	3.75	3.44	3.22	2.27	2.09	2.03	
% variance explained	11.72	10.74	10.07	7.11	6.53	6.35	

Note: *N* = 344; INC-RES = incongruity-resolution; NON = nonsense; SEX = sexual; DIS = disparagement; *h*² = communality. Expected loadings are italicized. Loadings ≥ .30 are listed.

Table 3
Intercorrelations among the EAHU scales and Cronbach's alphas

	INC-RES _f	NON _f	SEX _f	BLACK _f	M-DIS _f	W-DIS _f	INC-RES _a	NON _a	SEX _a	BLACK _a	M-DIS _a	W-DIS _a
INC-RES _f	<i>.74</i>	.42**	.46**	.27**	.44**	.36**	-.11*	-.09	-.04	-.03	-.05	-.01
	NON _f	<i>.83</i>	.40**	.28**	.35**	.29**	-.10	-.01	-.09	-.13*	-.11*	-.09
		SEX _f	<i>.85</i>	.42**	.49**	.50**	-.08	-.02	-.32**	-.11*	-.08	-.13*
			BLACK _f	<i>.81</i>	.45**	.50**	-.01	-.02	-.09	-.38**	-.05	-.18**
				M-DIS _f	<i>.72</i>	.45**	-.05	-.01	-.13*	-.09	-.38**	-.02
					W-DIS _f	<i>.76</i>	-.01	-.05	-.15**	-.22**	-.09	-.39**
						INC-RES _a	<i>.76</i>	.52**	.53**	.33**	.54**	.46**
							NON _a	<i>.89</i>	.41**	.04	.26**	.16**
								SEX _a	<i>.89</i>	.43**	.59**	.48**
									BLACK _a	<i>.84</i>	.45**	.55**
										M-DIS _a	<i>.75</i>	.49**
											W-DIS _a	<i>.82</i>

Note: N = 344 = p < .05; ** = p < .01

INC-RES = incongruity-resolution; NON = nonsense; SEX = sexual; M-DIS = man disparagement; W-DIS = woman disparagement; f = funniness. a = aversiveness. Cronbach's alphas are italicized in the diagonal.

Cronbach's alphas ranged from .72 (man disparagement humor) to .85 (sexual humor) in funniness scores, and from .75 (man disparagement humor) to .89 (nonsense humor and sexual humor) in aversiveness scores.

The average intercorrelation was r = .40 for funniness (coefficients ranged from r = .27 to r = .50). A positive correlation appeared between the funniness scores of the content categories and funniness of the two structural humor categories. Positive intercorrelations were found between the aversiveness scores of the six categories. The average intercorrelation was r = .42 (coefficients ranged from r = .04 to r = .59). Finally, the average intercorrelation between the funniness and aversiveness scores of each type of humor was r = -.14 (coefficients ranged from r = -.01 to r = -.39). These coefficients confirm the relative independence of the funniness and aversiveness scores in humor appreciation.

Discussion

The aims of this study are clear, and follow the basic structure of pilot studies of test items (Henrysson, 1971), and of studies aimed at obtaining evidence of content validity (Rubio, Berg-Weger, Tebb, Lee & Rauch, 2003). Therefore, the discussion of the results takes each of these objectives as a point of reference.

The first step taken involved the initial pool of 200 items and consisted of using judges to assess to what extent the items were conceptually appropriate for the facet they had been selected for theoretically. Judges help rule out items that are unclear, confusing or inappropriate from a theoretical point of view. This stage is considered to be key in the

process of creating an assessment tool. (Gordon, 2004; Smith, Fischer & Fister, 2003; Sireci, 1998).

The data obtained from the judges' evaluation led to eliminating 27 items in total. A qualitative estimation such as the one made in this study on item-facet match (Sireci, 1998) is necessary. However, this process cannot be considered sufficient to "back up" the proposed dimensionalization of the items or their fitness. Content validity analyses are just one step in the work aimed at generating an assessment tool, and the results obtained are insufficient. This is why there is a recommendation to obtain empirical evidence; the pre-pilot and pilot study of the items are the first steps in this direction (Mastaglia et al., 2003). Both studies were carried out with this aim; their most important findings are discussed below.

The discussion about the statistical analysis of the items should begin with a reflection about the sample used in the pre-pilot study. Given the high number of items, the group of 173 items had to be divided into four different booklets, each of which was dealt with by a different sample. Statistics calculated for item scoring are known to fluctuate depending on the sample involved. This makes it problematic to compare results of items that have been dealt with by different participants. In spite of this, we followed the procedure recommended for cases in which the item sample is so large that it is not advisable for one single subject to respond to each and every item. However, in these cases the samples used should be as homogeneous as possible, in order for the data to be comparable to a certain extent. In the study presented here, the samples came from the same assessment context, they were similar in size, and the different booklets were randomly assigned to the participants. In order to check the similarity of the item

scores of the participants from the different samples, an analysis of variance was carried out separately for each of the isolated dimensions of humor appreciation. No statistically significant differences were found.

The descriptive statistics of the selected elements from the pre-pilot study can be summarized into several points. First of all, the response range was always equal to or greater than 3. It was 4 in most of the cases. Second, the standard deviation of the items was always equal to or greater than 1 in the funniness response scale of all the dimensions, as well as in the aversiveness scale of the three content dimensions. As we pointed out in the section on statistical analyses, mean values were mostly low and close to 0 and featured little variability in the the aversiveness scale of the cognitive dimensions. As regards discrimination of the selected items, it should be noted that all the elements featured a corrected item-total correlation equal to or greater than .30. Most of the values were around .60. Besides, no item score differences were found to be positive or equal to 0 between the subjects situated within the higher or lower 27 per cent in the total score of the dimension. These metric properties of items were confirmed in the second study that formed the item analysis stage presented here. For the second study, we used a larger sample ($n = 344$) to work with a first proposal for a scale with 40 items selected from the pre-pilot stage (EAHU₄₀).

As for the variability found in the aversiveness scores of cognitive dimensions, it is worth noting the findings of the studies carried out by Willibald Ruch. In the theoretical approach of our dimensional proposal on humor appreciation, we adopted the cognitive dimensions used by Ruch. Ruch's review (1992) contained a table with all the studies carried out until the time it was published, which included the descriptive statistics of the 3WD – its assessment tool – in its different versions. The results show that, as regards funniness, the scores assessed by its scale in the three humor categories (incongruity-resolution, nonsense humor and sexual humor) were similar, as we found in our study. However, in the aversiveness scale, the scores of incongruity-resolution humor and nonsense humor were considerably lower than those obtained for the only content category Ruch dealt with – sexual humor.

Ruch (1992) cited the work of Ruch and Hehl (1986), where the mean range of item scores for the aversiveness scale was between .31 and 1.29 for incongruity-resolution, between .19 and 2.70 for nonsense humor, and between 1.22 and 2.96 for sexual humor. An analysis of the score distribution reveals (Ruch & Hehl, 1998) that in the structural categories of aversiveness, variance is low and values are around 0 (no aversiveness). All these results match the findings of our second item analysis. This second study confirmed the results of our pre-pilot analysis and corroborated the data on score distribution obtained by Ruch's group.

In the pre-pilot study, the homogeneity of the dimensions was studied by looking at the correlation of each item with the total value of its dimension, as well as the total value

of each of the other dimensions. The findings of this study can be summarized into two different statements: 1) every element always has a closer relation with the total value of its dimension than with any of the rest; however, 2) there are positive relations – which are in many cases significant – between items and dimensions other than their own.

The dimensional proposal this study is based on (Carretero-Dios et al, 2006) already suggested that the dimensions of humor appreciation are related to each other. Basically, it argues that any humor stimulus is related to a specific cognitive process, whether the latter has a clear specific content or not. Thus, there is a shared cognitive component, which must be reflected in the existence of a positive relation between the dimensions.

In spite of the postulated relation of the items with dimensions other than their own, it has been argued that, as far as possible, content should be isolated from structure (or cognitive processes) and the different structures and contents should be separated from each other (Carretero-Dios et al., 2006). We argue that, in order to clarify the variance provided by each dimension, as well as the specific relation of each dimension with other variables, distinctive facets should be distinguished as far as possible and common ones should not be mixed (Jackson, 1970). The dimension referring to a content should be understood mainly as content, because what the structure provides is delimited by the dimensions isolated for the structure itself. This is why our calculation procedure was aimed at obtaining a greater separation between each of the dimensions and the rest, through a greater homogenization of the distinctive features of each of them. These findings were confirmed in the study of the internal structure of the EAHU that followed the pre-pilot stage.

The empirical dimensionality of the EAHU₄₀ was analyzed with principal component analysis and Varimax rotation. The results showed the dimensions that had been suggested theoretically from the outset and their reliability was appropriate in every case. Nevertheless, the results led to dividing disparagement humor into two dimensions, man disparagement and woman disparagement humor. This finding has theoretical relevance, given that the theories that analyze disparagement humor (Ford, et al., 2008; Herzog 1999) highlight the importance of considering which sex is disparaged. The results obtained here empirically support these theories.

As regards the correlation analysis of the items with all the totals, on this occasion the data obtained for the aversiveness scale should be considered from a conceptual point of view. In this scale, the scores of the items of any of the content dimensions correlated more strongly – and in many cases significantly – with the aversiveness scores of the remaining content dimensions. In contrast, the correlations between the same content dimensions were uneven and moderate in the funniness response. The correlations of the structural dimensions with the aversiveness

scales were also considerably lower. It would be important to take these results into account when obtaining external evidence of validity, given that in some cases there may be common relations between aversiveness responses towards content facets and other variables. Moreover, the data highlight the importance of distinguishing humor responses with a positive character (funniness) from those with a negative valence (aversiveness). This point is dealt with in the discussion about obtaining validity evidence. After selecting the elements of each dimension, the funniness and aversiveness scores were correlated. We expected a negative and moderate relation between both responses, as was found in other studies (Ruch, 1992; Ruch & Hehl, 1986). This prediction was confirmed when the intercorrelation between the EAHU scales was explored.

We shall now deal with the aversiveness scale of the cognitive dimensions. Non-specific content items with a structure of incongruity-resolution or nonsense humor correlated more strongly – and in many cases significantly – with the aversiveness scale of the rest of the dimensions. Our comments about the score distribution of the aversiveness scale of the cognitive dimensions should be recalled here. The low standard deviation and mean values of this scale in the structural dimensions seem to have an influence on correlation data.

In short, the items selected to develop the EAHU are supported by the evidence shown about content validity, the metric properties of the items, the exploratory factor structure and the reliability values of each dimension. These results must be replicated using various samples of a greater size. Confirmatory factor analysis and multi-group confirmatory factor analysis should be applied in future research. Moreover, future studies will have to obtain external evidence of the validity of the EAHU scores (e.g., relationship between EAHU scores and sex, age or personality). Finally, our findings provide cross-cultural support to the results reported by Ruch and his group.

References

- AERA, APA & NCME. (1999). *Standards for educational and psychological tests*. Washington DC: American Psychological Association, American Educational Research Association, National Council on Measurement in Education.
- Beck, C.T., & Gable, R.K. (2001). Ensuring content validity: An illustration of the process. *Journal of Nursing Measurement, 9*, 201-215.
- Bing, J. (2007). Liberated jokes: sexual humor in all-female groups. *Humor: International Journal of Humor Research, 20*, 337-366.
- Bozikas, V.P., Kosmidis, M.H., Tonia, T., Garyfallos, G., Focas, K., & Karavatos, A. (2007). Humor Appreciation in Remitted Patients with Bipolar Disorder. *Journal of Nervous & Mental Disease, 195*, 773-775.
- Carretero-Dios, H., & Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales [Standards for the development and review of instrumental studies]. *International Journal of Clinical and Health Psychology, 5*, 521-551.
- Carretero-Dios, H., & Pérez, C. (2007). Standards for the development and review of instrumental studies: considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology, 7*, 863-882.
- Carretero-Dios, H., Pérez, C., & Buela-Casal, G. (2006). Dimensiones de la apreciación del humor [Dimensions of humor appreciation]. *Psicothema, 18*, 465-470.
- Cattell, R.B., & Luborsky L.B. (1947). Personality factors in response to humor. *Journal of Abnormal and Social Psychology, 42*, 402-421.
- Cattell, R.B., & Tollefson, D.L. (1966). *The IPAT humor test of personality*. Champaign, IL: Institute for Personality and Ability Testing.
- Clark, L.A., & Watson, D. (2003). Constructing validity: Basic issues in objective scale development. In A.E. Kazdin (ed.), *Methodological issues & strategies in clinical research (3rd ed.)* (pp. 207-231). Washington, D.C.: APA.
- Crocker, L., Llabre, M., & Miller, M.D. (1988). The Generalizability of Content Validity Ratings. *Journal of Educational Measurement, 25*, 287-299.
- Eysenck, H.J. (1943). An experimental analysis of five tests of "appreciation of humor". *Educational and Psychological Measurement, 3*, 191-214.
- Eyssel, F., & Bohner, G. (2007). The Rating of Sexist Humor Under Time Pressure as an Indicator of Spontaneous Sexist Attitudes. *Sex Roles, 57*, 651-660.
- Ford, T.E., Boxer, C.F., Armstrong, J., & Edel, J.R. (2008). More than Just a Joke: The Prejudice-Releasing Function of Sexist Humor. *Personality and Social Psychology Bulletin, 34*, 159-170.
- Galloway, G., & Chirico, D. (2008). Personality and humor appreciation: evidence of an association between trait neuroticism and preferences for structural features of humor. *Humor: International Journal of Humor Research, 21*, 129-142.
- Gordon, J. (2004). Developing and improving assessment instruments. *Assessment in Education: Principles, Policy and Practice, 11*, 243-245.
- Grant, J.S., & Davis, L.L. (1997). Selection and Use of Content Experts for Instrument Development. *Research in Nursing & Health, 20*, 269-274.
- Haladyma, T. (2004). *Developing and validating multiple-choice test items*. New York: Lawrence Erlbaum Associates.
- Haynes, S.N., Richard, D.C.S., & Kubany, E.S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment, 7*, 238-247.
- Henrysson, S. (1971). Gathering, analyzing and using data on test items. In R.L. Thorndike (ed.), *Educational Measurement (2nd ed.)* Washington, DC: American Council on Education.
- Henson, R., & Douglas, J. (2005). Test Construction for Cognitive Diagnosis. *Applied Psychological Measurement, 29*, 262-277.
- Herzog, T.R. (1999). Gender differences in humor appreciation revisited. *Humor: International Journal of Humor Research 12*, 411-423.

- Jackson, D.N. (1970). A sequential system for personality scale development. In C.D. Spielberger (ed.), *Current topics in clinical and community psychology* (vol. 2) (pp. 61-96). New York: Academic Press.
- Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison Wesley.
- Martínez-Arias, R. (1995). *Psicometría: teoría de los test psicológicos y educativos [Psychometry: theory of educational and psychological tests]*. Madrid: Síntesis.
- Mastaglia, B., Toye, C., & Kristjanson, L.J. (2003). Ensuring content validity in instrument development: Challenges and innovative approaches. *Contemporary Nurse* 14, 281-291.
- McGhee, P.E., Ruch, W., & Hehl, F.J. (1990). A personality-based model of humor development during adulthood. *Humor: International Journal of Humor Research*, 3, 119-146.
- Muñiz, J., Hidalgo, A.M., García-Cueto, E., Martínez, R., & Moreno, R. (2005). *Análisis de ítems [Item analysis]*. Madrid: La Muralla.
- Nunnally, J.C., & Bernstein, I.J. (1995). *Psychometric Theory*. New York: McGraw-Hill.
- Osterlind, S.J. (1989). *Constructing Test Items*. London: Kluwer Academic Publishers.
- Polit, D., Beck, C.T., & Owen, S. (2007). Is the CVI an acceptable indicator of content validity? *Research in Nursing & Health*, 30, 459-467.
- Rubio, D.M., Berg-Weger, M., Tebb, S.S., Lee, E.S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27, 94-104.
- Ruch, W. (1992). Assessment of appreciation of humor: Studies with the 3 WD humor test. In C.D. Spielberger and J.M. Butcher (eds.), *Advances in personality assessment*, (pp. 27-75). Hillsdale, NJ: Erlbaum.
- Ruch, W., & Hehl, F.J. (1986). Conservatism as a predictor of responses to humour-I. A comparison of four scales. *Personality and Individual Differences*, 7, 1-14.
- Ruch, W., & Hehl, F.J. (1998). A two-model of humor appreciation: Its relation to aesthetic appreciation and simplicity-complexity of personality. In W. Ruch (ed.), *The sense of humor: Explorations of a personality characteristic* (pp. 109-142). Berlin: Mouton de Gruyter.
- Samson, A.C., & Huber, O. (2007). The interaction of cartoonist's gender and formal features of cartoons. *Humor: International Journal of Humor Research*, 20, 1-25.
- Sireci, S.G. (1998). Gathering and analyzing content validity data. *Educational Assessment*, 5, 299-321.
- Smith, G.T., Fischer, S., & Fister, S.M. (2003). Incremental validity principles in test construction. *Psychological Assessment*, 15, 467-477.
- Spaan, M. (2006). Test and item specifications development. *Language Assessment Quarterly*, 3, 71-79.
- Wynd, C.A., Schmidt, B., & Atkins-Schaefer, M. (2003). Two quantitative approaches for estimating instrument content validity. *Western Journal of Nursing Research*, 25, 508-518.

Received May 3, 2007

Revision received October 3, 2008

Accepted October 13, 2008

APPENDIX 1

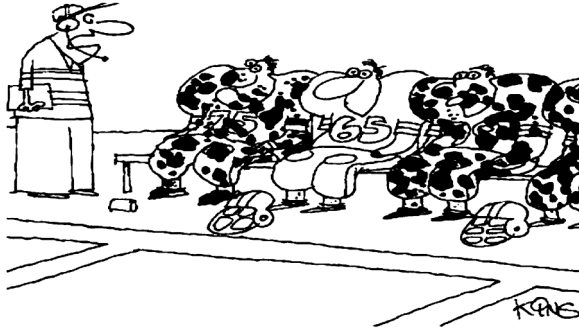
PROPOSAL FOR A SEMANTIC DEFINITION OF HUMOR APPRECIATION

<p>When faced with a joke, situation, etc., and given its internal structure, two basic cognitive processes may take place – incongruity-resolution and nonsense. The operative indicators of humor appreciation are obtained by crossing the subcomponents with these processes</p>		<p>Incongruity-resolution A fact which appears not to make sense finally becomes meaningful thanks to a logical rule</p>		<p>Nonsense A meaningless or ambiguous fact is presented and generates more incongruity or is partially resolved</p>	
		<p>A Situation, joke, or cartoon with no predominant content and a structure that leads to perceiving an incongruity that is eventually resolved</p>		<p>B Situation, joke, or cartoon with no predominant content and a structure that leads to perceiving an incongruity that is not resolved, is resolved partially, generates several solutions, etc.</p>	
<p>C Situation, joke, or cartoon with a predominant sexual content and involving an incongruity-resolution process</p>		<p>E Situation, joke, or cartoon with a predominant superiority-disparagement content and involving an incongruity-resolution process</p>		<p>F Situation, joke, or cartoon with a predominant superiority-disparagement content and involving an incongruity that is not resolved</p>	
		<p>G Situation, joke, or cartoon with a predominant black humor content and involving an incongruity-resolution process</p>		<p>H Situation, joke, or cartoon with a predominant black humor content and involving an incongruity that is not resolved</p>	
<p>CONSTRUCT ASSESSED Humor appreciation Dimension of sense of humor that refers to the degree of funniness and aversiveness shown towards certain humor stimuli depending on their content/subject and their internal structure (degree of coherence between the elements that form the humor stimuli)</p>	<p>COMPONENTS</p>	<p>SUBCOMPONENTS</p>	<p>With no specific content Group of jokes, cartoons, etc., with no specific subject</p>		
			<p>Sexual Jokes or cartoons on human and animal sexual life that deal with the various issues related to this subject</p>		
<p>Content Specific subjects used to group jokes, cartoons, humorous situations, etc.</p>	<p>COMPONENTS</p>	<p>SUBCOMPONENTS</p>	<p>Superiority-disparagement Jokes or cartoons that attempt to ridicule the physical, psychological, social or other characteristics of a given group or person, such as the sex they belong to, the region they are from, the job they do, etc.</p>		
			<p>Black humor Jokes or cartoons on dramatic events, catastrophes, major disasters, or physical or other damage generally suffered by humans</p>		

APPENDIX 2

EXAMPLE OF ITEMS FOR EACH DIMENSION OF HUMOR APPRECIATION

Incongruity-resolution humor (INC-RES)



"I don't want to name any names, but somebody is not putting 100% effort in the field"

Nonsense humor (NON)

A guy is walking in the street clapping his hands. Someone sees him and asks:

-What are you doing?

+I'm catching aureaflatoform godobrons.

-What are they like?

+I don't know, I haven't caught any yet!

Sexual humor (SEX)

-Mary, Mary!, bring me an orange, please

+Do you want me to peel it? [Translator's note: this can also mean "Do you want me to give you oral sex?"]

-OK!, and bring me the orange after that.

Black humor (BLACK)

¿Why did Adolf Hitler commit suicide?.....Because he couldn't pay the gas bill.

Disparagement humor (DIS)

Which is the longest part of a woman's body?.....The broomstick!

APPENDIX 3

INSTRUCTIONS FOR EXPERTS IN THE CONTENT VALIDITY TASK

You have been given a table specifying the components of the **humor appreciation** construct, as well as a booklet that includes 40 jokes or comic cartoons.

The table distinguishes between 8 possibilities, identified by the letters A, B, C, D, E, F, G and H. Each of these possibilities corresponds to one of the components of **humor appreciation**. Your task is to read each of the jokes or cartoons in the booklet and say which component of humor appreciation it belongs to. To do so, you will have to write the letter associated to the component of humor appreciation that best corresponds to the characteristics of the joke or cartoon according to you. See the example below:

Joke: Which is the hardest part of a vegetable?.....The wheelchair!

Classify this element filling the box with the letter of the component of humor appreciation that you think it belongs to

A

The letter “A” in the box means that, according to the table provided, the person who assessed this joke considered that this item belongs to the component of *humor with no specific content and with incongruity-resolution*.

To carry out this task, you should have the table that shows the components of **humor appreciation** in front of you. You may find it difficult to classify some jokes or cartoons into one of the 8 possibilities. In such cases, please say which option you believe is the most appropriate.

After each joke or cartoon, you will find a space for *observations*. Please use this space if you consider it is necessary to make any comments on the format (compression, presentation format, grammatical errors or other kinds of errors, etc.).

It is important to complete the task carefully. If you are tired or think your performance may not be good, it is better to stop and resume the task at some other moment.

At this point, we encourage you to start the task. **Thank you very much.**