



International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ORIGINAL ARTICLE

Content Validity through Expert Judgment for the Depression Clinical Evaluation Test

María Guillot-Valdés*, Alejandro Guillén-Riquelme, Gualberto Buela-Casal

Mind, Brain and Behavior Research Center, University of Granada, Spain

Received 15 November 2021; accepted 14 December 2021

Available online 4 February 2022



KEYWORDS

Depression
Expert judgment
Assessment
Content validity
Instrumental study

Abstract

Background/Objective: The evaluation of depression requires valid and reliable measuring instruments, which collect a wide spectrum of symptoms that this disorder displays, in order to carry out an accurate and differential diagnosis. The objective of this work is the construction of the Depression Clinical Evaluation Test (DCET), where affective, somatic, cognitive, behavioral and interpersonal symptoms are considered and also analyze its content validity through an expert judgment. **Method:** Based on different diagnostic and manual classifications, a specification table for a depression test was established. In its evaluation, 16 experts in Psychological Assessment, Psychometry and/or Psychopathology participated. A total of 300 items were created. The experts had to assess the items according to the criteria of Content, Relevance, Clarity, Comprehension, Sensitivity, and Offensiveness. In addition, 50 adults, evaluated the compression of the items. **Results:** The degree of understanding for all the items was high and the expert judgment favoured the suppression of 104 items, thus obtaining a shorter measuring instrument with a total of 196 items for ease of application. **Conclusions:** The content validity of the test is adequate and fits the agreed definition of depression.

© 2022 The Author(s). Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

PALABRAS CLAVE

Depresión;
Juicio de expertos;
Evaluación;
Validez de contenido;
Estudio instrumental

Validez de contenido mediante juicio de expertos para el Test de Evaluación Clínica de la Depresión

Resumen

Antecedentes: La evaluación de la depresión requiere de instrumentos de medida válidos, fiables y que recojan el amplio espectro de síntomas que este trastorno conlleva, para poder llevar a cabo un diagnóstico certero y diferencial. El objetivo de este trabajo es la construcción del Test de Evaluación Clínica de la Depresión (TECD), que contempla síntomas afectivos, somáticos, cognitivos, conductuales e interpersonales, y analizar su validez de contenido a través de un

* Corresponding author: María Guillot-Valdés, Mind, Brain and Behavior Research Center, University of Granada, Campus de Cartuja s/n, Granada, Spain

E-mail address: mguillot@ugr.es (M. Guillot-Valdés).

juicio de expertos. *Método:* A partir de diferentes clasificaciones diagnósticas y manuales se estableció la tabla de especificación del test para este cuestionario de depresión. En la evaluación de este participaron 16 expertos en Evaluación Psicológica, Psicometría y/o Psicopatología. Se crearon 300 ítems en total, que los expertos tuvieron que valorar atendiendo a los criterios de Contenido, Relevancia, Claridad, Comprensión, Sensitividad y Ofensividad. Además, 50 adultos, valoraron la comprensión de los ítems. *Resultados:* El grado de comprensión de todos los ítems fue elevado y el juicio de expertos supuso la supresión de 104 ítems, obteniendo así un instrumento de medida más breve, con 196 ítems en total, lo que facilitará su aplicación. *Conclusiones:* La validez de contenido del test es adecuada y se adapta a la definición de depresión establecida.

© 2022 The Author(s). Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Depression is one of the most common psychological disorders. According to World Health Organization (WHO) data, it is around 5.2% in the general population, being very close to those observed in other studies, around 7.2% (Lim et al., 2018). The study of depression has aroused interest over the years; and currently there has been a proliferation of work on the prevalence (Bueno-Notivol et al., 2021) and analysis of depression symptoms due to the COVID-19 health crisis (Cecchini et al., 2021). Incidentally, the evaluation of depression is complex even when there are a variety of instruments for its and diagnosis (Guillot-Valdés et al., 2019, 2020) and even in primary care with short evaluations (Rezaeizadeha et al., 2021).

A difficulty of depression assessment lies in the fact that it is a disorder with wide and varied symptomatology. This range includes cognitive, behavioural and psychosomatic symptoms in addition to the main emotional symptoms of the disorder. There are no scales on which all of them are evaluated with different items for each type of symptom. One of the most classic and used questionnaires is the Beck Depression Inventory (BDI-II; Beck et al., 1988). One of its advantages is that it covers a wide spectrum of depression with very few items; however, as mentioned above, it only covers each facet with one or two items. This fact makes it difficult to know the most affected areas of a specific case in a reliable way. Thus, it is common for evaluations to be complemented by using various specific questionnaires in order to make a reliable clinical profile of each affected area. This methodology presents results that are not easy to integrate and evaluate independent aspects of depression.

Last but not least, there is a controversy about whether depression has a dimensional or a categorical character, which prevails in current mental disorders classification systems. This approach influences the construction of instruments for the evaluation and diagnosis of the disorder (Chiesa et al., 2017). However, there are also contributions that emphasize the existence of an orthogonal structure between the two which would imply that obtaining high scores in positive affect would not lead to low scores in negative affect (Watson et al., 2011). Currently, very few questionnaires are focused on the dimensional approach to depression; therefore, the Basic Depression Questionnaire and the State/Trait Depression Inventory constitute certain examples on which some recent studies have been developed (Guillot-Valdés et al., 2019, 2020). Although they do not cover the entire symptom picture of depressive disorder.

The task of constructing a test implies careful planning, a clear and concrete vision of what it intends to measure, and that the items are well written and include a representative sample of the possible behaviours to be assessed (Muñiz et al., 2013; Muñiz & Fonseca-Pedrero, 2019). In this case, it is about operationalizing a construct, through concrete and tangible elements (items) (Carretero-Dios & Pérez, 2007; Muñiz & Fonseca-Pedrero, 2019). For this, a detailed process and a multitude of experts are required to help in the review of decisions. One of the most used methods to find the content validity of a questionnaire is the judgment of experts, who can either suggest which items the instrument should consist of to define the construct to be measured, or as in this case, evaluate the items already created based on a series of quantitative criteria (giving scores) or qualitative and suggesting, or adding any change to their wording if they consider it necessary (Garrote & del Carmen Rojas, 2015). This procedure is widely used by researchers to analyze the content validity of newly created instruments (Leyton-Román et al., 2021) or for adaptations of existing instruments (Cervilla et al., 2021).

The aim of this study is, first, establish a test specification table. For this we expect to establish an integral model that evaluates the main components of depression, thus covering all of the related symptoms. Secondly, we will develop an item bank test that cover this test specification table, including a proportional number of items for each factor and subfactor. The second aim is to estimate the content validity of this item pool, based on expert judgments of the Clinical Evaluation of Depression Test (TECD). In addition, it is intended to analyze the degree of understanding of the item bank to verify that they are intelligible to adult population.

Method

Participants

The sample, selected by convenience, consists of 16 experts and all of them had PhDs degrees in Psychology, with years of expertise and voluntarily agreed to participate in the study. They were specialized in the area of psychological evaluation, psychometry and/or clinical psychopathology and had great experience in the subject due to their academic training and work experience. Thus, they were able to provide adequate information, evidence,

judgments and evaluations (Escobar-Pérez & Cuervo-Martínez, 2008).

The criterion that different authors considered for the selection of judges was followed (Skjongs & Wentworth, 2001; Urrutia et al., 2014; Varela-Ruiz et al., 2012). Not only the already mentioned criterion but also the impartiality, motivation to participate, adaptability and availability of the judges were taken into account. They were contacted by e-mail, explaining the purpose of the project and requesting their collaboration.

In parallel and following the model of other authors (Fernández-Gómez et al., 2020; García-Cortés & Hernández, 2021; Luque-Vara et al., 2020) a pre-test of comprehension of the items was carried out, in which a total of 50 voluntarily collaborators participated ($M_{age} = 38$, $SD_{age} = 19.07$, 56% women) and to whom, as in the case of the experts, part of the questionnaire (50 items each) was also sent via email. Informed consent was obtained from each of them. The aspects to be evaluated were the degree of understanding of the item, reflected in the question as 'If the item was understood well' and the response ranged from *bad* (0) to *perfect* (10). The participants were also asked if there were any words that they did not understand and, finally, if they would express the item in another way and how. Subsequently, the mean of these scores was calculated to determine the degree of comprehension.

Instrument

For the creation of the Depression Clinical Evaluation Test (DCET) the 'Standards for educational and psychological testing' (American Educational Research Association, et al., 2014) and the guidelines of the International Test Commission (2016) were followed. In addition, several general articles on the creation and adaptation of tests were followed (Almanasreh et al., 2019).

In the first phase, from the documentary review carried out, a definition of depression was established: it was understood as a series of mood disorders characterized by having a common core symptomatology and that could vary in intensity, frequency or in the specific presence of symptoms among themselves. Derived from this definition and all the material consulted, the factors that composed it were established, collecting a logical grouping of the characteristics established in the manuals. Consequently, the symptoms were grouped into the following factors: affective, physiological/somatic, cognitive, behavioural and interpersonal. The number of symptoms considered in each one ranged from 3 to 8.

The symptoms' weights were established in accordance with whether they appeared in the DSM-5 and / or in the ICD-10 and 11, giving double weight to those that were collected in all classifications (Table 1). These weights were percentages ranging between 8 and 33%.

Once the weights of the factors and sub-factors were established, a confirmation of this phase was carried out by the experts. In addition, the most accurate response scale was utilized with respect to the proposed objective and it was presented to the experts as a 'table of test specifications' Two response scales were considered: one exclusively temporal with the evaluations marking the time of duration of the symptoms, and the other indicating the frequency of

appearance of the symptoms in three temporal moments (last month, last year and always). All the experts agreed that the best alternative was this second modality.

From there, only one change was proposed in the affective factor. Originally it was composed of depressed mood, anhedonia, and undervaluation and guilt each with a value of 33%, but after this initial trial depressed mood changed to 50% and anhedonia as well as undervaluation and guilt each became 25%.

After that, a bank of 300 items was prepared, where writing double negatives, double verbs, complex phrases and complex vocabulary was avoided. These items were subjected to qualitative evaluation by consulting six experts who were asked to indicate the adequacy of the definitions that were given of depression and each of the facets as well as the components that formed them. They were also asked to evaluate the sufficiency of the percentage of importance given to each facet in a component (established according to appearance in the DSM, the ICD or both).

Procedure

In the second phase, the second expert judgment coming from 13 judges (three of them also participated in the previous phase) was carried out. First, instructions were provided on the importance of this procedure and the tasks to be performed:

- 1) After the initial instructions, the general information of the test was presented so that the experts had all the necessary information to understand the complete final test and could provide their suggestions as to the general idea of the questionnaire and its objective.
- 2) Subsequently, the components and the facets of each of them were presented. Along with the definitions, the weight of the factor within the component was indicated (see Appendix A).
- 3) Then, the experts were asked to use the response scale.

In order to avoid the fatigue effect, the questionnaire was divided into six equal parts (50 items). Each of these parts had the same number of items for each factor and sub-factor (also disordered) to avoid both fatigue and response by acquiescence while trying to evaluate all the items of the same factor. Some experts were sent all parts of the questionnaire (300 items) and others only one (50 items) or two of them (100 items). In all cases, the criteria to be evaluated were the following:

- Content: the item belongs to the indicated factor and subfactor – No (0), Yes, just the factor (1), Yes (2).
- Relevance: the item is relevant to the construct – Not relevant (0), Needs some revision (1), Relevant, but minor revision (2), Relevant (3).
- Clarity: the item is clear or needs some revision – Confusing (0), Needs some revision (1), Slight revision (2), Clear (3).
- Comprehension: the item can be interpreted in different ways – No (0), In two ways (1), In several ways (2).
- Sensitivity: the item will allow differentiating between depression patients and subjects without the disorder – No, (0), In some cases (1), In most cases (2), Yes (3).

Table 1 Symptom summary for major depressive disorder / episode in DSM-5, ICD-10 and 11.

Symptom	DSM-5 Major depressive disorder	ICD-10 Major depressive episode	ICD-11 Depressive episode
Depressed mood	X	X	X
Loss of pleasure or interest in almost all activities	X	X	X
Significant weight gain	X		X
Significant weight loss	X	X	X
Significant increase in appetite	X		X
Significant loss of appetite	X	X	X
Insomnia	X	X	X
Hypersomnia	X	X	X
Psychomotor agitation	X		X
Psychomotor slowing down	X		X
Fatigue or loss of energy	X	X	X
Feeling worthless or excessive guilt	X	X	X
Decreased ability to concentrate	X	X	X
Decreased ability to think / make decisions	X		
Recurring thoughts of death	X		X
Suicidal plans or ideation	X	X	X
Suicide attempt	X	X	X
Reduced activity level		X	
Decreased attention		X	X
Loss of self-confidence		X	
Feeling of inferiority		X	
Grim perspective of the future		X	X
Self-harm		X	
Loss of reactivity to pleasant events and stimuli		X	X
Loss of libido		X	X

- Offensivity: the item may offend the evaluated persons – No (0), In some cases (1), In most cases (2), Yes (3).

The qualitative observations of the experts were considered for each of the items that formed the original instrument. In total, five judgments were obtained from each part into which the instrument was divided (50 items). Information was obtained from each of the experts individually (following the individual aggregate method) in a confidential manner, without them having contacted each other (Almenara & Cejudo, 2013). The data were collected in a Microsoft Excel 2010 sheet and then processed in the SPSS 25 statistical programme. This work was approved by the Ethics Committee of the University of Granada (Spain).

Results

All of the items that met the established requirements were considered adequate. Those that were partially adequate and required some changes and the inadequate ones that were considered incongruous or problematic with the established criteria were eliminated.

First, the adequacy of the item content – in this case depression – was analyzed to the measured construct. All those items with scores below 1.6 were eliminated (this

scale ranges from 0 to 2). Following this criterion, 39 items were eliminated (13% of the total).

Then, items with clarity less than 2.2 (scale from 0 to 3) were eliminated, thereby eliminating 17 items (6% of the total).

The next criterion was relevance, where items with a mean of less than 2.4 (scale from 0 to 3) between the five experts were taken as the cut-off point. When applying this criterion, the following 42 items were eliminated (14% of the total).

Finally, we observed the presence of items that, having acceptable scores, had various areas with scores that were not maximum and these items also exhibited slight comprehension problems. Here, 8 items were removed (4% of the total).

This process involved the suppression of 104 items. Some of the items (10) were corrected in writing. All this made it possible to obtain a clearer and slightly shorter measuring instrument, with 196 items, which helped to reduce the application time and improve the objectivity of the response options.

Table 2 shows the number of items that finally remained in each Factor and Subfactor.

In addition to the expert judgment, the 300 items were subjected to comprehension evaluation in an adult's sample. The responses of the 50 people surveyed were taken into

Table 2 Number of items corresponding to each Factor and Subfactor of the DCET.

Subfactor	Factor (primary)	Items n.
Depressed mood	Affective	25
Anhedonia	Affective	13
Undervaluation and guilt	Affective	5
Thoughts of undervaluation	Affective	7
Vacuum sensation	Somatic	1
Sleep disturbances	Somatic	8
Appetite / weight alteration	Somatic	4
Fatigue	Somatic	8
Motor agitation	Somatic	3
Language slowing	Somatic	2
Pain	Somatic	3
Decreased libido	Somatic	3
Disinterest in activities	Cognitive	8
Decreased concentration	Cognitive	8
Decreased attention	Cognitive	12
Thoughts of death	Cognitive	10
Expressions of discomfort	Behavioural	4
Abandonment of pleasant activities	Behavioural	7
Variation in diet	Behavioural	3
Worst task performance	Behavioural	8
Self-harm / suicide	Behavioural	9
Addictive substance abuse	Behavioural	3
social deterioration	Interpersonal	7
Family deterioration	Interpersonal	6
Work / school impairment	Interpersonal	7
Partner impairment	Interpersonal	6
Deterioration of other áreas	Interpersonal	6
Clinical discomfort	Interpersonal	10

account (scoring their understanding on a scale of 0 to 10) with an average comprehension of 9.82 out of 10. There were no items with an understanding lower than 9, which indicated that all the items were easily understandable and, therefore, it was not necessary to delete or modify any item after the analysis.

Discussion

The objective of this work was to propose a comprehensive model of depression in order to develop a test for its evaluation. Secondly, it was intended to estimate the content validity based on expert judgments of the DCET which

included five dimensions of the disorder for adults. Finally, the authors wanted to evaluate the comprehension of the developed items. After the different analyses, a test specification table was developed which adequately described the clinical criteria. From it, a sensitive and valid a bank of items was created, after purification. In addition, the items were understandable.

One of the strengths of this instrument is that it has been created with the intention of exhaustively evaluating those main, core and representative components of depression that are not present in cases of pure anxiety. This fact represents advancement over current questionnaires (e.g., BDI, Beck et al., 1988; CBD, Peñate, 2001; IDER, Spielberger et al., 2008). Likewise, it should be noted that the initial item bank that constituted the instrument was so exhaustive that the entire symptomatic picture of depressive disorder was covered as grouped by the following factors: affective, somatic, cognitive, behavioral and interpersonal. Also, various subfactors were considered within each one of them. This fact corresponds with the current psychometric specifications (Muñiz et al., 2013).

This work was submitted to an evaluation of its quality by experts. They evaluated them based on various categories (relevance, representativeness, etc.), thus making this procedure an essential criterion to determine the quality of measurement by an instrument (Muñiz & Fonseca-Pedrero, 2019). Incidentally, Almenara and Cejudo (2013) pointed out among the most outstanding benefits of this methodology, the level of depth it offered, the little difficulty one would experience using it or that the technical and human requirements for its utilization were not too demanding.

The present study selected 16 experts to respond to the proposed objectives, a number that was in the range recommended by various authors (Urrutia et al., 2014; Varela-Ruiz et al., 2012). Experts in the field of clinical psychology were selected and it was determined that all of them had to have experience in research and treatment on emotional as well as depressive disorders and psychometrics.

In view of the results obtained, one can have an instrument that has adequate content validity to evaluate depression and its symptoms. Furthermore, the sub-factors that compose them are also adjusted to the theoretical definition of depression proposed. This will be essential when evaluating depression comprehensively and will help them to know the main affected areas for the treatment (Mavranezouli et al., 2020; Pybis et al., 2017). Also, it is essential to have evaluation instruments with a dimensional and non-categorical approach. Currently, the ICD-11 (World Health Organization, 2019) recommends the use of these types of approaches as they can more appropriately address various disorders (e.g., personality disorders; Chiesa et al., 2017; Fowler et al., 2015; Waugh et al., 2017).

This work is not without its limitations. One of the most outstanding was the large number of items that the instrument initially covered, which meant dividing the questionnaire when presenting it to the experts. Considering future works in obtaining evidence of construct validity, future exploratory developments should also take into account maintaining an adequate number of items in each subfactor, taking special care in factors with few items. The choice of the number of experts was also somewhat difficult, due to

differences among the authors. Some considered the ideal range between 7 and 30 (Urrutia et al., 2014). Most authors recommend consulting more than 10 experts (García-Martín et al., 2016; Juárez-Hernández & Tobón, 2018). Thus, for the present study, altogether 16 experts were chosen (6 for the first phase and 10 for the second) for their availability as well as level of experience in the matter.

Future researches will be focused on applying pertinent statistical analyses (EFA, CFA) which allow selecting the items that will finally constitute each of the factors and sub-factors with adequate statistical significance. In any case, the authors of this work have managed to develop a pilot instrument to assess depression in a multidimensional way.

Funding

This study has been funded by Bursary FPU17/05262 for University Professor Training as part of the first author's thesis (Psychological Doctoral Programme B13 56 1; RD 99/2011).

Acknowledgment

The authors want to thank all the experts who have collaborated selflessly with them on the work so that it can meet its objectives as well as the pilot sample.

Appendix A Instructions for experts (presented in an Excel sheet)

As you know, within the process of creating a questionnaire, an expert judgment is necessary to guarantee the validity of the content and that each of the items is adequate and representative of the construct. Mentioned below are 50 items belonging to a questionnaire for the evaluation of depression, which in turn consists of 300 items.

You have to evaluate each item in terms of its content, relevance, clarity, comprehension, sensitivity and offensiveness, answering in each of the columns provided (Excel sheet). There is also a section for comments, where you can add any suggestions or clarifications. The items have been divided into factors:

- I. **AFFECTIVE SYMPTOMS:** This dimension consists of those emotional responses that occur with greater frequency or intensity in people with depressive disorders. It would include:
 1. **DEPRESSIVE MOOD:** feelings of sadness, discomfort and hopelessness towards the future. In adolescents and some adults irritability may manifest. Weight in the component: 50%.
 2. **ANHEDONIA:** inability to experience pleasure, loss of interest or satisfaction in almost all activities. Component weight: 25%.
 3. **UNDERVALUATION AND GUILT:** feelings of guilt, responsibility for adversity or illness, feelings of incapacity and mistrust towards oneself. Component weight: 25%.
- II. **SOMATIC SYMPTOMS:** This dimension is made up of those physical and bodily responses that are felt with greater frequency or intensity in people with depressive disorders.
 1. **SENSATION OF EMPTINESS:** feeling of emptiness, nerves, closed stomach. Component weight: 14%.
 2. **SLEEP DISORDERS:** difficulty in falling asleep, awakening with difficulty in returning to sleep, early awakening, daytime sleepiness. Sometimes displaying hypersomnia (more than 10 hours or more than 2 hours than baseline). Component weight: 14%.
 3. **ALTERATION OF APPETITE / WEIGHT:** marked increase or decrease in appetite. Significant increase or decrease in weight without diets (5% compared to the initial weight). Component weight: 14%.
 4. **FATIGUE:** feelings of tiredness, lack of energy and vitality. Tiredness from actions that previously did not cause fatigue. Muscular weakness. Component weight: 14%.
 5. **MOTOR AGITATION:** unable to stand still, tremors, repetitive movements, playing with small objects, clothes or the body itself, not being able to sit. Component weight: 14%.
 6. **SLOWING OF THE LANGUAGE:** slower speech, slower tone of voice, increased response latency, decreased vocabulary, mutism. Component weight: 14%.
 7. **PAIN:** pain in the joints or abdomen. Headaches. Component weight: 8%.
 8. **DECREASE OF LIBIDO:** decreased sexual desire, anorgasmia in women, erectile dysfunction. Component weight: 8%.
- III. **COGNITIVE SYMPTOMS:** This dimension includes those thoughts and ideas that are felt with greater frequency or intensity by people with depressive disorders. In the same way, the decrease in cognitive abilities (attention, concentration and problem solving) is manifest.
 1. **DISINTEREST IN ALMOST ALL ACTIVITIES:** inability to imagine a better future, thoughts of disinterest in pleasant activities. Component weight: 20%.
 2. **DECREASE IN CONCENTRATION:** slowing down of thinking, difficulty in concentrating and making decisions. It may imply a lack of memory. Component weight: 10%.
 3. **DECREASE IN ATTENTION:** difficulty in maintaining attention. Getting distracted easily. Component weight: 10%.
 4. **UNDERVALUATION THOUGHTS:** Thoughts of guilt, recalling past mistakes, misinterpreting everyday events as evidence of worthlessness, exaggeration of failures, unrealistic evaluations of one's worth / dignity, increases in self-criticism. Component weight: 20%.
 5. **THOUGHTS OF DEATH:** recurring thoughts of death, wishing to die, suicidal ideation, suicidal planning. Thoughts of self-harm. Believing that others would be better off if one died and wishing not to wake up. Component weight: 20%.
- IV. **BEHAVIOURAL SYMPTOMS:** This dimension is made up of those responses, actions or observable behaviours that are performed with greater frequency or intensity by people with depressive disorders.
 1. **EXPRESSIONS OF DISCOMFORT:** crying, making complaints, self-reproach for being ill or not achieving goals. Component weight: 18%.

2. ABANDONING PLEASANT ACTIVITIES: not doing pleasant tasks as before, doing them less frequently, and not really getting involved in activities. Component weight: 18%.
 3. VARIATION IN DIET: eating less, eating more, eating more carbohydrates and sweets, striving to eat as before. Component weight: 10%.
 4. WORST PERFORMANCE IN TASKS: worse performance in habitual tasks, employing a lot of effort to carry out the tasks, striving to carry out the same activity as before, decreased activity. Component weight: 18%.
 5. SELF-AGGRESSION / SUICIDE: self-harm, suicide planning, purchase of materials, setting a time and place for suicide, making a suicide attempt. Component weight: 18%.
 6. SUBSTANCE ABUSE: abuse of addictive substances, most commonly alcohol. Component weight: 18%
- V. INTERPERSONAL SYMPTOMS: This dimension is formed by those consequences in social relationships and work obligations that can occur due to suffering from a depressive disorder.
1. SOCIAL IMPAIRMENT: less time spent on social relationships or experiencing less enjoyment with them. Less pleasant social interactions. Isolation. Sensitivity to interpersonal rejection. Component weight: 12.5%.
 2. FAMILY IMPAIRMENT: less time spent with the family or experiencing less enjoyment with it. Component weight: 12.5%.
 3. WORK / SCHOOL IMPAIRMENT: difficulty in performing tasks properly in such a way as to entail a remarkable decrease noticed by bosses / teachers or colleagues. Component weight: 12.5%.
 4. COUPLE RELATIONSHIP IMPAIRMENT: less time spent with the spouse or experiencing less enjoyment with him/her. Reduction of sexual desire. Component weight: 12.5%.
 5. DETERIORATION IN OTHER AREAS: less time dedicated to other areas important to the person. Component weight: 20%.
 6. CLINICAL DISCOMFORT: self-perceived discomfort that makes the patient seek help to resolve this issue. Component weight: 30%.

References

- Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy, 15*, 214-221. <https://doi.org/10.1016/j.sapharm.2018.03.066>.
- Almenara, J. C., & Cejudo, M. D. C. L. (2013). La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y comunicación (TIC) [Application of expert judgment as a technique for evaluating information and communication technologies]. *Revista Eduweb, 7*, 11-22.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Beck, A., Epstein, N., Brown, G., & Steer, R. (1988). An inventory for measuring clinical anxiety. Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 893-897. <https://doi.org/10.1037/0022-006X.56.6.893>.
- Bueno-Notivol, J., Gracia-García, P., Olaya, B., Lasheras, I., López-Antón, R., & Santabárbara, J. (2021). Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies. *International Journal of Clinical and Health Psychology, 21*, Article 100196. <https://doi.org/10.1016/j.ijchp.2020.07.007>.
- 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.
- Cecchini, J. A., Carriedo, A., Fernández-Río, J., Méndez-Giménez, A., González, C., Sánchez-Martínez, B., & Rodríguez-González, P. (2021). A longitudinal study on depressive symptoms and physical activity during the Spanish lockdown. *International Journal of Clinical and Health Psychology, 21*, Article 100200. <https://doi.org/10.1016/j.ijchp.2020.09.001>.
- Cervilla, O., Vallejo-Medina, P., Gómez-Berrocal, C., & Sierra, J. C. (2021). Development of the Spanish short version of Negative Attitudes Toward Masturbation Inventory. *International Journal of Clinical and Health Psychology, 21*, Article 100222. <https://doi.org/10.1016/j.ijchp.2021.100222>.
- Chiesa, M., Cirasola, A., Williams, R., Nassisi, V., & Fonagy, P. (2017). Categorical and dimensional approaches in the evaluation of the relationship between attachment and personality disorders: An empirical study. *Attachment & Human Development, 19*, 151-169. <https://doi.org/10.1080/14616734.2016.1261915>.
- Escobar-Pérez, J., & Cuervo-Martínez, A. (2008). Content validity and expert judgment: An approach to its use. *Avances en Medicina, 6*, 27-36.
- Fernández-Gómez, E., Martín-Salvador, A., Luque-Vara, T., Sánchez-Ojeda, M. A., Navarro-Prado, S., & Enrique-Mirón, C. (2020). Content validation through expert judgement of an instrument on the nutritional knowledge, beliefs, and habits of pregnant women. *Nutrients, 12*, 1136. <https://doi.org/10.3390/nu12041136>.
- Fowler, J. C., Sharp, C., Kalpakci, A., Madan, A., Clapp, J., Allen, J. G., Frueh, B. C., & Oldham, J. M. (2015). A dimensional approach to assessing personality functioning: Examining personality trait domains utilizing DSM-IV personality disorder criteria. *Comprehensive Psychiatry, 56*, 75-84. <https://doi.org/10.1016/j.comppsy.2014.09.001>.
- García-Cortés, G. E., & Hernández, L. G. J. (2021). Validation of instruments to evaluate the educational model and degree of progress according to the knowledge society. *Atenas, 3*, 21-37.
- García-Martín, A., Antúnez, A., & Ibáñez, S. J. (2016). Analysis of the training process in expert players: instrument validation. *Revista Internacional de Medicina y Ciencias de la Actividad Física del Deporte, 16*, 157-181. <https://doi.org/10.15366/rimcafd2016.61.012>.
- Garrote, P. R., & del Carmen Rojas, M. (2015). La validación por juicio de expertos: dos investigaciones cualitativas en Lingüística aplicada. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas, 18*, 124-139. <https://doi.org/10.26378/mlael918259>.
- Guillot-Valdés, M., Guillén-Riquelme, A., & Buela-Casal, G. (2019). Reliability and validity of the Basic Depression Questionnaire. *International Journal of Clinical and Health Psychology, 19*, 243-250. <https://doi.org/10.1016/j.ijchp.2019.07.002>.
- Guillot-Valdés, M., Guillén-Riquelme, A., & Buela-Casal, G. (2020). A Meta-Analysis of the Generalization of the Reliability of State/Trait Depression Inventory Scores. *Psicothema, 32*, 476-489. <https://doi.org/10.7334/psicothema2020.106>.
- International Test Commission. (2016). The international test commission guidelines on the security of tests, examinations, and other assessments: international test commission (ITC). *International Journal of Testing, 16*, 181-204. <https://doi.org/10.1080/15305058.2015.1111221>.

- Juárez-Hernández, L. G., & Tobón, S. (2018). Analysis of the implicit elements in the content validation of a research instrument. *Revista Espacios*, 39, 23-28.
- Leyton-Román, M., Mesquita, S., & Jiménez-Castuera, R. (2021). Validation of the Spanish Healthy Lifestyle Questionnaire. *International Journal of Clinical and Health Psychology*, 21, Article 100228. <https://doi.org/10.1016/j.ijchp.2021.100228>.
- Lim, G. Y., Tam, W. W., Lu, Y., Ho, C. S., Zhang, M. W., & Ho, R. C. (2018). Prevalence of Depression in the Community from 30 Countries between 1994 and 2014. *Scientific Reports*, 8, Article 2861. <https://doi.org/10.1038/s41598-018-21243-x>.
- Luque-Vara, T., Linares-Manrique, M., Fernández-Gómez, E., Martín-Salvador, A., Sánchez-Ojeda, M. A., & Enrique-Mirón, C. (2020). Content Validation of an Instrument for the Assessment of School Teachers' Levels of Knowledge of Diabetes through Expert Judgment. *International Journal of Environmental Research and Public Health*, 17, Article 8605. <https://doi.org/10.3390/ijerph17228605>.
- Mavranzouli, I., Megnin-Viggars, O., Grey, N., Bhutani, G., Leach, J., Daly, C., Welton, N. J., Katona, C., El-Leithy, S., Greenberg, N., Stockton, S., & Pilling, S (2020). Cost-effectiveness of psychological treatments for post-traumatic stress disorder in adults. *PLoS One*, 15, Article e0232245. <https://doi.org/10.1371/journal.pone.0232245>.
- Muñoz, J., Elosua, P., & Hambleton, R. K. (2013). Guidelines for the translation and adaptation of tests: second edition. *Psicothema*, 25, 151-157. <https://doi.org/10.7334/psicothema2013.24>.
- Muñoz, J., & Fonseca-Pedrero, E. (2019). Ten steps for the construction of a test. *Psicothema*, 31, 7-16. <https://doi.org/10.7334/psicothema2018.291>.
- Peñate, W. (2001). Presentation of a basic questionnaire to evaluate the genuine symptoms of depression. Introduction. *Análisis y Modificación de Conducta*, 27, 679-731.
- Pybis, J., Saxon, D., Hill, A., & Barkham, M. (2017). The comparative effectiveness and efficiency of cognitive behaviour therapy and generic counselling in the treatment of depression: evidence from the 2 nd UK National Audit of psychological therapies. *BMC Psychiatry*, 17, 1-13. <https://doi.org/10.1186/s12888-017-1370-7>.
- Rezaeizadeh, A., Sanchez, K., Zolfaghari, K., & Madiad, N. D. (2021). Depression screening and treatment among uninsured populations in Primary Care. *International Journal of Clinical and Health Psychology*, 21, Article 100241. <https://doi.org/10.1016/j.ijchp.2021.100241>.
- Skjong, R., & Wentworth, B. (2001). *Expert Judgement and risk perception*. Stavanger: International Society of Offshore and Polar Engineers.
- Spielberger, C. D., Buela-Casal, G., & Agudelo, D. (2008). *Inventario de Depresión Estado/Rasgo (IDER)*. TEA Ediciones.
- Urrutia, M., Barrios, S., Gutierrez, N., & Mayorga, M. (2014). Optimal methods for determining content validity. *Educación Médica Superior*, 28, 547-558.
- Varela-Ruiz, M., Díaz-Bravo, L., & García-Durán, R. (2012). Description and uses of the Delphi technique in research in health areas. *Investigación Educación Médica*, 1, 90-95.
- Watson, D., Clark, L. A., & Stasik, S. (2011). Emotions and the emotional disorders: A quantitative hierarchical perspective. *International Journal of Clinical and Health Psychology*, 11, 429-442.
- Waugh, M. H., Hopwood, C. J., Krueger, R. F., Morey, L. C., Pincus, A. L., & Wright, A. G. (2017). Psychological assessment with the DSM-5 Alternative Model for Personality Disorders: Tradition and innovation. *Professional Psychology: Research and Practice*, 48, 79-89. <https://doi.org/10.1037/pro0000071>.
- World Health Organization. (2019). ICD-11 for mortality and morbidity statistics. <https://www.who.int/classifications/icd/en/>